

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 31, 2003, 16:22:15 ; Search time 2303 Seconds
(without alignments)
12257.374 Million cell updates/sec

Title: US-10-069-541-5
Perfect score: 1743
Sequence: 1 atgctttccatggaagg.....ctgaagataattacagtga 1743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estlin:*

4: em_estmd:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_estl:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	518.8	29.8	672	17	AG157499 Pan trogl
2	404	23.2	516	10	BE233479 139685 MA
3	329.8	18.9	650	10	BB626260 BB626260
4	312.8	17.9	541	10	AW668962 111664 MA
5	274.2	15.7	524	12	BE723927 198406 MA
6	263.2	15.1	800	9	AL669749 AL669749

7	212.4	12.2	1037	9	AL666817	AL666817
8	209.2	12.0	641	13	BI630566	BI630566 RH59836.5
9	207.4	11.9	640	13	BI629504	BI629504 RH59836.5
10	205.2	11.8	658	13	BM629925	BM629925 170006875
11	203.2	11.7	652	10	BB626456	BB626456 BB626456
12	186.4	10.7	624	13	BJ122485	BJ122485 BJ122485
13	179.8	10.3	580	13	BJ125564	BJ125564 BJ125564
14	166	9.5	604	9	AU199794	AU199794 AU199794
15	163.4	9.4	500	10	AV994375	AV994375 AV994375
16	163.2	9.4	555	13	BJ117801	BJ117801 BJ117801
17	155	8.9	561	17	AQ16435	AQ16435 RPI-11-10
18	153	8.8	584	13	BJ105382	BJ105382 BJ105382
19	134	7.7	632	17	AZ612750	AZ612750 BM0439117
20	128.2	7.4	500	13	BJ105730	BJ105730 BJ105730
21	127.6	7.3	618	17	AZ908709	AZ908709 RPI-24-2
22	121	6.9	926	17	CNS041317	AL295624 Tetraodon
23	117.2	6.7	523	13	BI508286	BI508286 BI170004A
24	117.2	6.7	530	13	BI503332	BI503332 BI170001A
25	112.8	6.5	355	9	AU209671	AU209671 AU209671
26	112.4	6.4	558	13	BI507950	BI507950 BI170001A
27	108	6.2	420	13	BI506529	BI506529 BI170027B
28	99.4	6.1	539	13	BI508323	BI508323 BI170007A
29	99.4	5.7	486	12	BG353180	BG353180 PS23603.Y
30	94.6	5.4	563	13	BI503188	BI503188 BI170001A
31	91.6	5.3	1058	17	CNS0586F	AL350160 Tetraodon
32	91.4	5.2	1101	17	CNS058BW	AL350357 Tetraodon
33	90.4	5.2	627	10	BB663292	BB663292 BB663292
34	84.4	4.8	936	17	CNS03COM	AL237199 Tetraodon
35	81.4	4.7	741	13	BJ153899	BI153899 BI153899
36	77.8	4.5	367	9	AL839758	AL839758 AL839758
37	77.6	4.5	292	13	BI502965	BI502965 BI170004A
38	76.6	4.4	760	17	CNS033GD	AL226102 Tetraodon
39	73.2	4.2	962	17	CNS033GC	AL226101 Tetraodon
40	73	4.2	311	9	AL839681	AL839681 AL839681
41	68.2	3.9	696	17	CNS03MCS	AL250597 Tetraodon
42	67.4	3.9	807	17	CNS024XK	AL181361 Tetraodon
43	66	3.8	360	14	CI3840	CI3840 CI3840 fuji
44	64.6	3.7	1083	17	CNS051WV	AL317416 Tetraodon
45	64.2	3.7	984	17	CNS03COL	AL237198 Tetraodon

ALIGNMENTS

RESULT 1
AG157499 672 bp DNA linear GSS 09-JAN-2002
LOCUS Pan troglodytes DNA, clone: RP43-022H02.T7, genomic survey
DEFINITION sequence.
ACCESSION AG157499.1 GI:16687177
KEYWORDS Pan troglodytes male lymphocytes DNA, clone_lib:RPI-43 Chimpanzee
SOURCE Male BAC Library clone:RP43-022H02.T7.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library RPI-43
TITLE Unpublished
REFERENCE 2 (bases 1 to 672)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
TITLE Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
JOURNAL (E-mail: chimpanzee@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
REFERENCE 3
Clones are derived from the chimpanzee BAC library RPI-43 This BAC end was generated during the R&D process and may have higher chance

of clone tracking errors.

PRIMERS

Sequencing: T7
LIBRARY

Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.

Location/Qualifiers

1..672

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="RP43-022H02.17"

/sex="male"

/cell_type="lymphocytes"

/clone_lib="RPC1-43 Chimpanzee Male BAC Library"

186 a 144 c 135 g 206 t 1 others

BASE COUNT

ORIGIN

Query Match 29.8%; Score 518.8; DB 17; Length 672;

Best Local Similarity 97.3%; Pred. No. 1.9e-134;

Matches 549; Conservative 0; Mismatches 12; Indels 3; Gaps 2;

1110 AATGCTTCGACAAAGAAATCGTTGGTTATCGGAATCACAGTGTGTGGTGGAGC 1169

Db 111 ACAGCTTCGACAAAGAAATCGTTGGTTATCGGAATCACAGTGTGTGGTGGAGC 170

QY 1170 ATCTGCAACAGCCATGGCCCTGCTGACGAAACCTGTATGGGCTCTGGTACCTCAGTTC 1229

Db 171 ATCTGCAACAGCCATGGCCCTGCTGACGAAACCTGTATGGGCTCTGGTACCTCAGTTC 1229

QY 1230 TGACCTGTTTACATGCTTATCTTCCGCCAGCTGCTTGTGTACTCTTGTAAAGGAAC 1289

Db 231 TGACCTGTTTACATGCTTATCTTCCGCCAGCTGCTTGTGTACTCTTGTAAAGGAAC 1289

QY 1290 CAACACCTATGGGCGCGTGGCAGTTATGTTCTGGCCCTTCTCTGAGATAACTGGAGG 1349

Db 291 CAACACCTATGGGCGCGTGGCAGTTATGTTCTGGCCCTTCTCTGAGATAACTGGAGG 1349

QY 1350 GGAGCCATATCTGATCTTACGCCCTTGATCTTACCCCTGGCTATACCCCTGATGATATA 1409

Db 351 GGAGCCATATCTGATCTTACGCCCTTGATCTTACCCCTGGCTATACCCCTGATGATATA 1409

QY 1410 TGGTATATATAATCAGAAATTTCCATTTAAACACTTGGCCATGGTGTACATCTTAAAC 1469

Db 411 TGGTATATATAATCAGAAATTTCCATTTAAACACTTGGCCATGGTGTACATCTTAAAC 1469

QY 1470 CAACATTTGCATCTCCTATCTAGCAAGATATCTATTG-AAAGTGGAACCTTGGCCACCTA 1528

Db 471 CAACATTTGCATCTCCTATCTAGCAAGATATCTATTG-AAAGTGGAACCTTGGCCACCTA 1528

QY 1529 AATTAGATGATTTGATGCTGTGTGCAAGACACAGTGAAGAAACATGGATAAGACAA 1588

Db 531 AATTAGATGATTTGATGCTGTGTGCAAGACACAGTGAAGAAACATGGATAAGACAA 1588

QY 1589 TTCTTGTCAAAATGAAATATTAATTAAGATGAACCTTGGCACTTGTGAAGCCACGACAGA 1648

Db 591 TTCTTGTCAAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAAT 1648

QY 1649 GCATGACCCCTCAGCTCACTTCA 1672

Db 649 ACATGACCCCTCAGCTCACTTCA 672

RESULT 2

BE233479

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 516)

Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,

Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.

and Keeler,J.W.

Design and use of two pooled tissue normalized cDNA libraries for

EST discovery in swine

Unpublished (2000)

Contact: Smith TEL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred

v0.980904.e. Vector identified by cross_match with the -minscore 18

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCTCCAGTCACGACG

Plate: 75 row: G column: 12

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

1..516

/organism="Sus scrofa"

/db_xref="taxon:9823"

/clone_lib="MARC 1PIG"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;

Library made from pooled tissue from day 11, 13, 15, 20,

and 30 embryos."

BASE COUNT 128 a 129 c 112 g 147 t

ORIGIN

Query Match 23.2%; Score 404; DB 10; Length 516;

Best Local Similarity 86.4%; Pred. No. 3.1e-102;

Matches 446; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 1067 GTTCCATGTTTCGACGGAACATCTACCACTTTCCTTCAGACAAAATGCTTCGGACAAAG 1126

Db 1 GTTCTATGTTTCTAGAAAACATCTAGCAGCTCTCATTCAGACAAAACGCTTCGACAGG 60

QY 1127 AAATCGTTGGTTATGCGAATCACAGTGTGTGTGTTGAGCAGCTGCAACAGCCATGG 1186

Db 61 AGATGCTCTGGGTCTATGGGATCACAGTATTCGTGTGTTGGTGGCTGCAACAGCCATGG 120

QY 1187 CTTCTGCTGACGAAACTGTGTATGGGCTGTGTACCTCAGTTCGACCTTGTTCATATCG 1246

Db 121 CTTCTGCTGACGAAACTGTGTATGGGCTGTGTACCTCAGTTCGACCTTGTTCATATCG 1246

QY 1247 TTATCTTCCCGCAGCTGCTTGTGTACTTGTGTTAAGGGAACCAACACCTATGGGGCGG 1306

Db 181 TTATCTTCCCGCAGCTGCTTGTGTACTTGTGTTAAGGGAACCAACACCTATGGGGCGG 1306

QY 1307 TGGCAGGTATGTTTCTGCGCTCTTCCTGAGAATAACTGGAGGGAGCCATATCTGTATC 1366

Db 241 TGGCAGGTATGTTTCTGCGCTCTTCCTGAGAATAACTGGAGGGAGCCATATCTGTATC 1366

QY 1367 TTCAGCCCTTGTATCTTCTACCTGCTATTTACCTGATGATAATGGTATATATATATCA 1426

Db 301 TGCAGCCCTTGTATCTTCTACCTGCTATTTACCTGATGATAATGGTATATATATCA 1426

QY 1427 AATTTCATTTAAACACTTGGCCATGTTTACATCATCTTTAAACCAACATTTGCACTTCCT 1486

Db 361 GATTTCATTTAAACACTTGGCCATGTTTACATCATCTTTAAACCAACATTTGCACTTCCT 1486

QY 1487 ATCTAGCCCAAGTATCTATTGGAAGTGGAACTTGGCCACCTAAATAGATGATTTGATG 1546

Db 421 ATCTAGCCCAAGTATCTATTGGAAGTGGAACTTGGCCACCTAAATAGATGATTTGATG 1546

QY 1547 CTGTGTTTCAAGACACAGTGAAGAAACATGGATA 1582

Tue Apr 1 13:57:25 2003

BI629504 640 bp mRNA linear EST 10-SEP-2001
 LOCUS RH58381.5prime RH Drosophila melanogaster normalized Head pFlc-1
 DEFINITION Drosophila melanogaster cDNA clone RH58381 5 similar to CG7708:
 FBan0007708 GO:[] located on: 3R 91B8-91B8; : 08/23/2001, mRNA
 sequence.

ACCESSION BI629504.1 GI:15531714

VERSION EST

KEYWORDS fruit fly.

SOURCE Drosophila melanogaster

ORGANISM Drosophila melanogaster

REFERENCE 1 (bases 1 to 640)

AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson

, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George

, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S.,

Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,

Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin

, G.M.

BDGP/HMI RH Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit genomic AE003723: arm:3R [14365708,14621331]

estimated-cyto:91B5-91D2: 08/23/2001

Plate: RH.583 row: G column: 9

High quality sequence stop: 512.

Location/Qualifiers

1..640

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RH58381"

/pFlc-1"

/sex="male and female"

/dev_stage="Adult"

/lab_host="DH5-alpha Tona"

/note="organ: head; Vector: pFlc1; Site:1: XhoI; Site:2:

BamHI; Library was kindly generated by Piero Carninci at

the RIKEN. The library was normalized and excised using

Cre recombinase. Plasmid cDNA library."

113 a 193 c 175 g 159 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 11.9%; Score 207.4; DB 13; Length 640;

Mismatches 256; Indels 3; Gaps 1;

Matches 374; Conservative 0;

769 TACTTTCAGAGGTTCTCTCTTCTCTCAGCCACCTATGCTCAAGTGTCTCTCTCTG 828

4 TACTTCAGCGGTGCTGCTCCAGCAACCGCAGGAGGCCAGCTTCTGCTCTATGTT 63

829 GCAGCTTTCGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 888

64 GCAGCGCCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 123

889 TCACAGACTTGCAGACAGTGCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 948

124 GCTACACTTGGACAGACAGATACAGAGGACCTATCCCTGACCGTGGAGAGACG 183

949 GACATGATTTACAAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1008

184 AGCATGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 243

1009 CTTGGTGCAGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1068

244 TTGGCGCGTGTTCCTGCGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 303

QY 1069 TCCATGTTTCCACGGAACACTCTACCAGCTTTCCTTCAGACAAATGCTTCGGAAGAAGAA 1128
 Db 304 TCCATGTTTCCACGGAACACTCTACCAGCTTTCCTTCAGACAAATGCTTCGGAAGAAGAA 363
 QY 1129 ATCGTTTGGGTATGCGGAATCACAGTCTTTGTGTTTGGAGCATCTGCAACAGCCATGGCC 1188
 Db 364 ATCATTTGGGTATGCGGAGTCGCCATCATTTGTTGGTGGCATCTTCCTGGCTTACCATGGCC 423
 QY 1189 TTGCTGACGAAACTGTGTATGGGCTCTGGTACCTCTGACCTTGTGTTTACATCGTT 1248
 Db 424 CTCACCATTCCTCCATCTACGGTTTGTGGTCCATGCTCGGATCTGCTTACCTCAT 483
 QY 1249 ATCTTCCCCAG---CTGCTTGTGTACTCTTTTAAAGGAACCAACACCTATGGGCC 1305
 Db 484 CTGTTCCCGCAGCTACTGATGTTGTCACCTTCAAGAAAGCACTGCAACACGCTACGCGCAG 543
 QY 1306 GTGCGAGTTATGTTTCTGGGCTCTCTCTGAGATTAACCTGAGGGAGCCATATCTGTAT 1365
 Db 544 CTGTGGCATACATTTGTGGCTTGGCCATCGGACTGTGCGGGCGGTGAGGCCATCTTGGGA 603
 QY 1366 CTTACGCTTGTATCTTCTACCTCGCTATTATAC 1398
 Db 604 CTGGCTCCATTGATCAAGTATCCCGCTACGAC 636

RESULT 10
 BM629925 658 bp mRNA linear EST 26-FEB-2002
 LOCUS 17000687500663 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
 DEFINITION 19600449651480 5', mRNA sequence.

ACCESSION BM629925
 VERSION BM629925.1 GI:18929436
 KEYWORDS EST
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae

REFERENCE 1 (bases 1 to 658)
 AUTHORS Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab
 , R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
 TITLE Celera Anopheles gambiae EST project
 JOURNAL Unpublished (2002)
 COMMENT Contact: Holt R.A.
 Celera Genomics
 45 W. Gude Dr., Rockville, MD 20850, USA
 Tel: 2404533151
 Fax: 2404534580
 Email: HoltRA@celera.com
 Plate: NU010049UI row: D column: 06
 Seq primer: M13 Reverse.

FEATURES
 source
 1..658
 /organism="Anopheles gambiae"
 /strain="RSP-ST (Reduced susc. to Permethrin - std.
 chromosome)"
 /db_xref="taxon:7165"
 /clone_lib="A.Gam.ad.cDNA1"
 /dev_stage="Adult"
 /lab_host="DH10b"
 /note="vector: pSport1; Site:1: SalI; Site:2: NotI; Whole
 adult mosquitoes (mixed sex) frozen on liquid nitrogen.
 cDNA inserts >500 bp cloned directionally into pSport 1.
 Not 1 site is 3'. Clones available through the Malaria
 Research and Reference Reagent Resource Center
 (www.malaria.mr4.org)."

BASE COUNT 106 a 199 c 186 g 167 t

ORIGIN
 Query Match 11.8%; Score 205.2; DB 13; Length 658;
 Best Local Similarity 59.1%; Pred. No. 2.3e-46;

QY 769 TACTTTCAGAGGTTCTCTCTTCTCTCAGCCACCTATGCTCAAGTGTCTCTCTG 828
 Db 4 TACTTCAGCGGTGCTGCTCCAGCAACCGCAGGAGGCCAGCTTCTGCTCTATGTT 63
 QY 829 GCAGCTTTCGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 888
 Db 64 GCAGCGCCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 123
 QY 889 TCACAGACTTGCAGACAGTGCATATGCTGCTGCTGCTGCTGCTGCTGCTGCT 948
 Db 124 GCTACACTTGGACAGACAGATACAGAGGACCTATCCCTGACCGTGGAGAGACG 183
 QY 949 GACATGATTTACAAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1008
 Db 184 AGCATGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 243
 QY 1009 CTTGGTGCAGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1068
 Db 244 TTGGCGCGTGTTCCTGCGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 303

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Prepare full-length cDNA libraries for cap-trapper-selected cDNAs to
genes. *Genome Res.* 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, I., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuyama,
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10
(11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y., and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)
Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
Hayashizaki, Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. *Mamm. Genome* 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

```
e mouse tissues.
Location/Qualifiers
1. .652
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="9330188K24"
/clone_lig="RIKEN full-length enriched, adult male
diencephalon"
/sex="male"
/tissue_type="diencephalon"
/dev_stage="adult"
/lab_host="DH10R"
```

/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using reverse transcriptase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCCAGTATTAATTAATCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:

[illegible]

```

Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGGATCCAGAGACTCTTTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATCTCGATTAAATTAATCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified plucapsript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
BamHI"

BASE COUNT      140 a      168 g      148 t      3 others
ORIGIN
Query Match      11.7%; Score 203.2; DB 10; Length 652;
Best Local Similarity 86.1%; Pred. No. 8.3e-46;
Matches 223; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 ATGCGTTTCCCATGCGGAAGGACTGATAGCTATCATCGTGTCTACCTCTCTAATTTGCTG 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 394 ATGTCFTTTCCACGTAGAAGGACTGGTAGCTATTATCCTCTTCTACTCTCTTATATTCTG 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 GTTGGAAATATGGCTGCTGGAGAACCCAAAACAGTGGCAGCGCAGAGAGCGCAGAA 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

454 GTTGAATATGGGTGTCATGGTAAACCAAAACACGCGCAACCCAGAGAGCGCAGTGAA 513
121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTGGTTGGTGGATTTACCATGACAGCT 180
122 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
514 GCCATCATAGTCGGGGCCGTCACATGGTTTGGTTGGTTGGTTGGTTTACCATGACAGCC 573
181 ACCTGGGTCGAGGAGGAGGTATCAATGGCAGCTCAAGCAGGTTTATGTACCAAGGTTAT 240
182 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
574 ACCTGGTTGGAGGAGGCTACATCAATGGACAGCAGACAGTCTATGTGGCCAGGTTGT 633
241 GGCCTAGCTTGGGCTCAGG 259
183 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
634 GGTCTAGCTTGGGCTCANG 652

RESULT 12
BJ122485 624 bp mRNA linear EST 23-JAN-2002
LOCUS BJ122485 unpublished oligo-capped cDNA library, C. elegans L1 stag
DEFINITION Caenorhabditis elegans cDNA clone yk1288e07 5', mRNA sequence.
ACCESSION BJ122485
VERSION BJ122485.1 GI:18282624
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 624)
Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.,
and Sugano, S.
A complementary view of the C.elegans genome
Unpublished (2002)
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. .624
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk1288e07"
/clone.lib="unpublished oligo-capped cDNA library, C.
elegans L1 stage"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"
BASE COUNT 153 a 129 c 162 g 180 t
ORIGIN
Query Match 10.7%; Score 186.4; DB 13; Length 624;
Best Local Similarity 60.2%; Pred. No. 4.3e-41;
Matches 353; Conservative
Qy 19 GGACTGATAGTCATCGTTCCTACCTCTCTAATTTTGTGGTTGGTAATATGGGTGCC 78
Db 45 GGTATCGTGGCCATGTGTTCTTCTAGTGCCTCTTCTGTGGTAATATGGCGGGT 104
Qy 79 TGGAGAACCAAAACAGCTGGCAGC-----GCAGAGACCGCAGCGAACCCATCA 127
Db 105 AGAAATCGAAAGTTCAAGAGCTTGAATCAGAGCGCGCGCGACGGAAGAGGTGA 164
Qy 128 TAGTTGGTGGCGGAGATATTGGTTATTGGTTGGTGGATTTACCATGACAGCTACCTGGG 187
Db 165 TGTTAGCTGGGAGAAACATCGGAACCTGTGTGCGAATTTTCAATGACTGCCACGTGGG 224

```

[illegible]

FEATURES	SOURCE
Location/Qualifiers	
1. 580	
/organism="Caenorhabditis elegans"	
/strain="N2"	
/db_xref="taxon:6239"	
/clone="wkl324f12"	
/clone_lib="unpublished oligo-capped cDNA library, C."	
/elegant_lib="stage"	
/sex="hermaphrodite"	
/tissue_type="whole animal"	
/dev_stage="L1"	
/note="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into ppc86"	
130 a	122 c 154 g 174 t
BASE COUNT	
ORIGIN	

[illegible]

```

Query Match      10.3%; Score 179.8; DB 13; Length 580;
Best Local Similarity 61.9%; Pred. No. 3e-39;
Matches 305; Conservative 0; Mismatches 182; Indels 6; Gaps 1;

QY 152 TATTGCTGTGGTACATGACAGCTACCTGGGTCGAGGAGGTATATCAATGGCA 211
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 TGTGTGCTTACTGTTTCAAAATGACTGCCACGTGGTGGCGGTCTATATCAATGGAA 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 212 CAGCTGAAGCAGTTTATACAGAGTTATGGCTAGCTTGGCTCAGGACCAATGGAT 271
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 CCGCCGAGGCTCTGTATATGAGGT-----CTCCTTGGATGTCAGGCTCCAGTTGGAT 115
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 272 ATTCTCTTACTGCTGATTTTAGTGGCTGCTTTTGCAGAACTATGCTTCAAAAGGGT 331
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116 ATGCAATTTCCCTTGTATGGGAGACTACTTTTCGCAAGAAATGCGAGAGAGAT 175
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 332 ATGTGACCATGTAGACCCGTTTACAGAAATCTATGAAACGCGATGGCGGACTCCCTGT 391
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176 ATATTACAAATGCTCATCTTTTCAGCAAAATATGGCCACGAATCGGTGGCTTGTATGT 235
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
392 TTATTCTGCACTGATGGGAGAAATGTTCTGGGCTGCGAGCAATTTTCTCTGCTTGGGAG 451
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
236 ATGTTCCAGCACTTCTTGGTGAACATCTTGAGACAGCAGCCATCTTTTCGSCACTTGGTG 295
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 452 CCACATCAGCGTGATCATCGATGATGACATTTCTGTCTATCATCTCTGCACCTCA 511
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
296 CAACACTGTGCGTAATCTTGGAAATCGACATGAATCATCAGTACCCTGTGCGGCTGTA 355
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 512 TTGCCACTCTGTACACACTGTTGGAGGGCTCTATTCTGTGGCTACACTGATGCTGCTTC 571
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
356 TTGCGGTATTTACACACTTCCCGGTGGATACTATGAGTCGCGGTACACTGACGTGCTGTC 415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 572 AGCTCTTTTGCATTTTGTAGGCTGTGGATCAGCGTCCCTTGGCATTTGCACTGCTG 631
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
416 AACTATTGTCATTTCTGCTGCTTGGGTTGCGTTCGCTGCGGCGGCTATGGTGCATGATG 475
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 632 CAGTCGCACACAT 644
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 476 GTCGAAGGATAT 488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
AUI99794
LOCUS
DEFINITION
AUI99794 unpublished oligo-capped cDNA library, stage L1
AUI99794
AUI99794
AUI99794.1 GI:14827195
EST.
Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
and Sugano, S.
A complementary view of the C.elegans genome
Unpublished (2001)
Contact: Yuji Kohara
Genome Biology Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. .604
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk769e12"
/clone_lib="unpublished oligo-capped cDNA library, stage
L1"

FEATURES
source

```

```

/sex="Hermaphrodite"
/dev_stage="L1"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"
BASE COUNT      151 a      125 c      158 g      170 t
ORIGIN

Query Match      9.5%; Score 166; DB 9; Length 604;
Best Local Similarity 58.8%; Pred. No. 2.3e-35;
Matches 333; Conservative 0; Mismatches 215; Indels 18; Gaps 2;

QY 19 GGACTGATAGCTATCATCGTGTTCACCTTCTAATTTTCTGCTGGTGGAAATATGGCTGCC 78
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45 GTATCGTGGCCATTTGTTCTTCTACGTGCTCAATCTTGTCTGTTGGAATATGGCGGT 104
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 79 TGGAGAACCAAAA-----ACAGTGGCAGCGCAGAGAGCGGAGCAAGCCATC 126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105 AGAAATCGAAAGTTCAAAAGAGCTTGAATCAGAGCCGCGCGGCGAGAGAGGTG 164
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 127 ATAGTGTGTCGCCAGATATTTGTTTATGTTGGTGGATTTTACCATGACAGCTACCTGG 186
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165 ATGTTAGCTGGAGAAACATCGGAACTCTTTCGGAATTTTTCACAAATGACTGCCAGTGG 224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 187 GTCGAGGAGGGTATATCAATGCGACAGCTGAAGCAGTTTATGTACAGGTTATGGCCTA 246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
225 GTTGGCGGTCTTATATCAATGGAACCGCGAGCTCTGTATATGAGGT-----CTC 278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 247 GCTTGGGCTCAGCACCACCAATTGATATCTCTTAGTCTGATTTTATGCTGCTTCTT 306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 CTGATGTCAGGCTCCAGTTGGATATGCAATTTCCCTTGTATGGGAGGACTACTTTTC 338
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 307 GCAAAACCTATGCTTCAAAGGGGTATGTGACCATCTTACACCGTTTCAGCAATCTAT 366
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 339 GCAAGAAATCGAGAGAGATATATACAACTGCTCGATCTTTTCAGCAAAATAT 398
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 367 GGAACACGATGGCGGACTCTCTGTTTATTCGTCACCTGATGGGAAATGTTCTGGGCT 426
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 399 GCCCAAGCAATCGTGGCTGTATGTATGTCAGCACTTCTTGGTGAACATTTCTGGACA 458
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 427 GCAGCAATTTCTCTGCTTTGGGAGCCACCATCAGGCTGATCATCGATGATGATGAC 486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 459 GCAGCAATTTCTGCGCACTTGGTGCACACTGTCGGTAATCTTGGAAATCGACATGAAT 518
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 487 ATTTCTGTCATCATCTCTGCACTCATTTGCCACTCTGTACACACTGTTGGGAGGCTCTAT 546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 519 GCATGAGCACTCTGCGGCTGTATGTCGCTATTTCTACACATTCACCGGTGATACTAT 578
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 547 TCTGTGGCTACACTGATGCTGTTCA 572
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 579 GCAGTCGCGTACACTGAGTCTGTTCA 604
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
AUI994375
LOCUS
DEFINITION
AUI994375 Nori Satoh unpublished cDNA library, EST 15-MAR-2002
AUI994375
AUI994375
AUI994375.1 GI:19485709
EST.
Clona intestinalis.
Clona intestinalis.
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Clona.
Satoh, N., Satoh, Y., Kohara, Y. and Shin-I, T.
Expressed genes in Clona intestinalis
Unpublished (2000)
Contact: Nori Satoh

```


Tue Apr 1 13:57:25 2003

Department of Zoology
Kyoto University
Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoheascidian.zool.kyoto-u.ac.jp.

FEATURES
source
1. .500
Location/Qualifiers
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="citb37j13"
/clone_lib="Nori Satoh unpublished cDNA library, tailbud
embryo"
/tissue_type="whole animal"
/dev_stage="tailbud embryo"
/note="Vector: pBluescript SK"

BASE COUNT 110 a 109 c 107 g 173 t 1 others
ORIGIN

Query Match 9.48; Score 163.4; DB 10; Length 500;
Best Local Similarity 62.6%; Pred. No. 1.2e-34;
Matches 271; Conservative 0; Mismatches 161; Indels 1; Gaps 1;
QY 845 TGGTATGGCCATCCAGCCATCTCAATGGGGCCATTCAGACATCAACAG-ACCTGGAAAC 903
Db 1 TGTTCATGTCATACCTTCGATATTAAATCGGTGCAATTCGTCATCTACAGNATTTGAAAC 60
QY 904 CAGATGTCATATGGCTTCAGATCCCAAGACTACAGAGAGGCGACAGATGATTTAGCA 963
Db 61 GCAACATCGTAGGGCTCCCAAGTGTGAAAGGGGCGACCAAGCCAAATATCTACCC 120
QY 964 ATTTGTTGTCAGTATCTGCGCTGTGTATATTTCTTTTGGTCTGGTGCAGTTTCT 1023
Db 121 ATTTGCTTCATACCTCACCCCTGTAGCTGTATCATCTTTGGGCTTGGCGCTGTTCT 180
QY 1024 GCTGCTGTATGTCATCAGCAGATCTTCCATCTGTCAGCAAGTTCCATGTTTGCACGG 1083
Db 181 GCTGCTGTATGTCATCAGCAGATCTTCCATCTGTCAGCAAGTTCCATGTTTGCACGG 240
QY 1084 AACATCTACAGCTTTCCCTTCAGACAAATGCTTCGACAAAGAAATCGTTGGGTATG 1143
Db 241 AATATTTATACCTCGTCATCCGACCAAGGCATCGGAGTTGGAGCTGTGTGGGTAATG 300
QY 1144 CGAATCACAGTGTGTGTTTGGAGCATCTGCAACGCCATGCGCTTGTGACGAAAACT 1203
Db 301 AGAGTATCAATAATATTCATAGCGGAGGAGCCACTGCTTTAGCTCTTGTGTGAATCT 360
QY 1204 GTGTATGGCTCTGGTACCTCAGTCTGTGACCTTCTTTACATCGTTATCTTCCCCAGCTG 1263
Db 361 GTTATTTATCTTTCTATCTTTGTTGGACCTGATCTACGTCATCTTTTCCACAGTTT 420
QY 1264 CTTTGTGCTACTCT 1276
Db 421 GTAAGTGTCTCT 433

Search completed: March 31, 2003, 18:28:17
Job time : 2317 secs

GenCore version 5.1.4.p5.4578
Copyright (C) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2003, 16:26:35 ; Search time 194 Seconds
(without alignments)
7645.597 Million cell updates/sec

Title: US-10-069-541-5
Perfect score: 1743
Sequence: 1 atgcttcacatggaagg.....ctgaagataattacagtga 1743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 574371 seqs, 425486471 residues 1148742
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	242.6	13.9	1461	10	US-09-974-300-501
2	155	8.9	455	10	US-09-864-761-1838
3	72	4.1	96	10	US-09-864-761-18589
4	41	2.4	2839	10	US-09-995-007-1
5	39.6	2.3	2028	10	US-09-733-630-1
6	39.6	2.3	2456	10	US-09-733-630-3
7	38.2	2.2	2028	9	US-09-928-530-3
8	38.2	2.2	2028	9	US-10-162-012-28
9	38.2	2.2	2326	9	US-09-928-530-1
10	38.2	2.2	2326	9	US-10-162-012-26
11	36.6	2.1	148567	10	US-10-254-869-3
12	36.6	2.1	148567	10	US-09-801-876B-3
13	35.6	2.0	5629	9	US-10-092-154-1162
14	35.6	2.0	5629	10	US-09-764-847-1162
15	35.6	2.0	6265	9	US-10-092-154-1161
16	35.6	2.0	6265	10	US-09-764-847-1161
17	35.4	2.0	234	9	US-10-046-935-1856
18	35.4	2.0	234	9	US-09-878-178-1856
19	35.4	2.0	314	10	US-09-983-965-359

c	20	35.2	2.0	173808	12	US-10-003-806-10	Sequence 10, Appl
	21	35	2.0	344	10	US-09-960-352-13002	Sequence 13002, A
	22	34.8	2.0	1776	10	US-09-815-242-9527	Sequence 9527, Ap
c	23	34.8	2.0	2712	10	US-09-974-300-1927	Sequence 1927, Ap
c	24	34.8	2.0	12482	9	US-10-270-877-25	Sequence 25, Appl
c	25	34.8	2.0	12482	9	US-10-270-837-25	Sequence 25, Appl
	26	34.2	2.0	10996	10	US-09-070-927A-202	Sequence 202, App
	27	34.2	2.0	368004	10	US-09-949-654-3	Sequence 3, Appl1
	28	34	2.0	683	9	US-09-984-245-17	Sequence 17, Appl
	29	34	2.0	683	9	US-09-966-262-17	Sequence 17, Appl
c	30	34	2.0	1109	7	US-08-781-986A-222	Sequence 22, App
c	31	34	2.0	465237	10	US-09-933-267A-1	Sequence 1, Appl1
c	32	33.8	1.9	1791	9	US-10-119-988-9	Sequence 4803, Ap
c	33	33.8	1.9	1993	9	US-10-119-988-9	Sequence 9, Appl1
c	34	33.8	1.9	3178	10	US-09-954-456-1882	Sequence 1882, Ap
c	35	33.8	1.9	3178	10	US-09-969-347-189	Sequence 189, App
c	36	33.8	1.9	176373	9	US-10-095-407-17	Sequence 17, Appl
c	37	33.8	1.9	1985	10	US-09-764-870-26	Sequence 26, Appl
c	38	33.6	1.9	2857	9	US-09-978-295A-263	Sequence 263, App
c	39	33.6	1.9	2857	9	US-09-978-697-263	Sequence 263, App
c	40	33.6	1.9	2857	9	US-09-978-192A-263	Sequence 120, App
c	41	33.6	1.9	2857	9	US-10-066-500-120	Sequence 263, App
c	42	33.6	1.9	2857	9	US-09-999-832A-263	Sequence 263, App
c	43	33.6	1.9	2857	9	US-09-978-189-263	Sequence 105, App
c	44	33.6	1.9	2857	9	US-10-174-590-105	
c	45	33.6	1.9	2857	9		

ALIGNMENTS

RESULT 1
US-09-974-300-501
; Sequence 501, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 501
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-501

Query Match	13.9%	Score 242.6;	DB 10;	Length 1461;
Best Local Similarity	52.4%	Pred. No. 3e-62;		
Matches	663;	Conservative	0;	Mismatches 554;
				Indels 48;
				Gaps 4;
Qy	151	TTATTGTTGGTGGATTACCATGACAGCTACCTGGGTGGGAGGAGGTATATCAATGGC	210	
Db	125	TTTTTCGCGAATGGTGACGATGCGCGCAACATGGGTGGCGCGGATATATTACGGA	184	
Qy	211	ACAGCTGAAGCAGTTTATGTACCAAGTTATGGCTTAGCTTCAGGCACCAATTCGA	270	
Db	185	ACGCCGAATGACCTTACA-----GCGACGGCCTCATCTGGGCGCAAGCGCTTGGGGC	238	
Qy	271	TATTCTCTTACTCTGATTTAGGTGGCTGTCTTTCGAAACCTATTCGGTTCACAAAGGG	330	
Db	239	TACGATTGACCGGATTATCGCGGTATTTCTTCGCCAGAAATCGCGGTATCAAA	298	
Qy	331	TATGTGACCATGTTAGACCCCGTTTCAGCAAAATCTATGAAACACGATGGCGGACTCTG	390	

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.97
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
OTHER INFORMATION: NT HIT: gill141884, EVALDE 5.00e-33
US-09-864-761-18589

Query Match 4.1%; Score 72; DB 10; Length 96;
Best Local Similarity 100.0%; Pred. No. 9.7e-12;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 824 TCCTGGCAGCTTTCGGGTGCTGGTATGCCCATCCAGCCATCACTATGGGGCCATTG 883
Db 1 TCCTGGCAGCTTTCGGGTGCTGGTATGCCCATCCAGCCATCACTATGGGGCCATTG 60
QY 884 GAGCATCAACAG 895
Db 61 GAGCATCAACAG 72

RESULT 4
US-09-995-007-1
Sequence 1, Application US/09995007
Patent No. US20020123102A1
GENERAL INFORMATION:
APPLICANT: NANCY CARRASCO, ET AL.
TITLE OF INVENTION: THYROID SODIUM/IODIDE SYMPORTER AND NUCLEIC ACID ENCODING SAME
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/995,007
FILING DATE: 26-NO. US20020123102A1-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/595,553A
FILING DATE: FEBRUARY 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: CRAIG J. ARNOLD
REGISTRATION NUMBER: 34,287
REFERENCE/DOCKET NUMBER: 96700/393
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2839
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: <unknown>
DESCRIPTION: OLIGONUCLEOTIDE
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: RAT
INDIVIDUAL ISOLATE: SODIUM/IODIDE SYMPORTER
SEQUENCE DESCRIPTION: SEQ ID NO: 1;

QY 741 GATGCTGGTGAATCCCATGCAAGCATACTTTTCAGAGGGTTCTCTTCTTCTCCTCAGC 800
Db 266 GATGCTGGTGAATCCCATGCAAGCATACTTTTCAGAGGGTTCTCTTCTTCTCCTCAGC 325
QY 801 CACCTATGCTCAAGTCTGCTCTTCTTCTGCGAGCTTTCGGGTGCTGGTATGCCCATCC 860
Db 326 CACCTATGCTCAAGTCTGCTCTTCTTCTGCGAGCTTTCGGGTGCTGGTATGCCCATCC 385
QY 861 AGCCATCACTATGGGGCCATGGAGCATCAACAG 895
Db 386 AGCCATCACTATGGGGCCATGGAGCATCAACAG 420

RESULT 3
US-09-864-761-18589
Sequence 18589, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Rank, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 18589
LENGTH: 96
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC009963.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2


```

Db 303 TGAACCTAATGGCTGTTTCTGTGCTGATGTTGGCCCTGGATCTCTCCCTACCCATCTACAT 362
QY 318 GCGTTCAAAGGGGTATGTGACCATGTTAGACCGCTTTCAGCAAAATCTATGGAAACGCAT 377
Db 363 TGCTGGTCAGTCACACGATGCCAGATACCTACGGAAGCCGCTTCGGTGGCATCAGAAAT 422
QY 378 GGGGGGACTCGTGTATTCCTGCACTGATGGGAAATGTTCTGGGCTGCAGCAATTTT 437
Db 423 CCCCATCATCTGCTGCTACTCTACCTATTATCTACATCTTCACCAAGATCTCGTAGA 482
QY 438 CTCTGCTTTGGGAGCCACCATCAGCGTCATCATCG--ATGGATATGCATTTCTGT 494
Db 483 CATGATGCAAGTGCCATCTCTTCCACGACGCTTTCGACACTGTTGGGAGGCGCTATTCTGTGCG 554
QY 495 CATCATCTCTGCACACTATTGCCACTCTGTACACACTGGTGGGAGGCGCTATTCTGTGCG 593
Db 543 AGTTGGGCTACTGGCCATCACTGCTGTATACACGGTTGCTGGTGGCGCTGCTGTGTAT 602
QY 555 CTACACTGATGTCGTTTCAGCTCTTTTGCATTTTGTAGG 593
Db 603 CTACACGGATGCCCTGCAGACGCTGATCATGCTTATAGG 641

RESULT 9
US-09-928-530-1
; Sequence 1, Application US/09928530
; Patent No. US20020156002A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Silos-Santiago, Inmaculada
; TITLE OF INVENTION: 32620, A NOVEL HUMAN SODIUM-SUGAR
; TITLE OF INVENTION: SYMPORTER FAMILY MEMBER AND USES THEREOF
; FILE REFERENCE: 10446-080001
; CURRENT APPLICATION NUMBER: US/09/928,530
; CURRENT FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/227,068
; PRIOR FILING DATE: 2000-08-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2326
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (178)...(2202)
US-09-928-530-1

```

Db 600 CCCCATCATCTCTGGCTGTACTCTACCTATTATCTACATCTTACCAAGATCTCGGTAGA 659
 QY 438 CTCTGCTTTGGGAGCCACCATCATGCGTATCG---ATGTGGATATGCACATTTCTCTG 494
 Db 660 CATGATGAGGTTGCCATCTTTCATCCAGCAGCTTTCCGCACCTGGATCTGTACTCGCCAT 719
 QY 495 CATCATCTCTGCACCTCATTTGCCACTCTGTACACACTGGTGGAGGGCTCTATTCTGTGGC 554
 Db 720 AGTTGGGCTACTGGCCATCACTGCTGTATACACGGTTGCTGGTGGCCTGGCTGCTGTAT 779
 QY 555 CTACACTGATGTCGTTACAGCTCTTTTGCATTTTGTAGG 593
 Db 780 CTACACGGATGCCCTGCACAGCTGATCATGCTTTATAGG 818

```

RESULT 10
US-10-162-012-26
; Sequence 26, Application US/10162012
; Publication No. US20030051660A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Silos-Santiago, Inmaculada
; APPLICANT: Gu, Wei
; TITLE OF INVENTION: NOVEL HUMAN ION CHANNEL AND TRANSPORTER FAMILY MEMBERS
; FILE REFERENCE: 10448-190001
; CURRENT APPLICATION NUMBER: US/10/162,012
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 60/209,845
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: US 09/875,321
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US01/18340
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/209,257
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,423
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18398
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/209,238
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,363
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18247
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/227,068
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 09/928,530
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/25475
; PRIOR FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: US 60/226,770
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/934,421
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26096
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/279,281
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 10/109,029
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: PCT/US02/09728
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/290,288
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US (not assigned)
; PRIOR FILING DATE: 2002-05-13
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 2326
; TYPE: DNA

```

Matches 123; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

1416 ATATAATCAGAAATTTCCATTTTAAACACACTTCCCATGGTTACATCATCTTAAACCAACAT 1475
 31673 AGATAATCAGTGTGTTTAACTTTTAAATTAAGCAGTAGGAGAAATGACTTTTGGGAACCTT 31732
 1476 TTGCATCTCCTATCTAGCCCAAGTATCTATTGAAAGTGGAACTTGGCCACCTAAATTTAGA 1535
 31733 AGAAATTTGGAAACCTTTTATTTCTATGATTGAATATCAACTATATGTAATTTAGTCTTAAG 31792
 1536 TGATTTTGTGCTGTTTGTGCAAGACACAGTGAAGAAACATGGATAAGACAATTTCTTGT 1595
 31793 GTTATATGCTAGAAACATTTTCAAAAACGAAAGCAGCAGCAATGACATCAAAAATGCAATGT 31852
 1596 CAAAATGAAATATTAATTAATAGATGAATGCACTTGTGAAGCCAGCAGACAGATGAC 1655
 31853 CAAAAGCAATGGTTTAAATAGAAATACATCATTTTAAACAATCTTGAAGTTTAAAGAT 31912
 1656 CCTCAGCTCAACTTTTCCCAATAAAGA 1682
 31913 CCTATAAAATCACAAACCCAGAAGGA 31939

RESULT 12
 US-09-801-876B-3
 ; Sequence 3, Application US/09801876B
 ; Patent No. US20020127683A1
 ; GENERAL INFORMATION:
 ; APPLICANT: YE, Jane et al
 ; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
 ; FILE REFERENCE: CL001160
 ; CURRENT APPLICATION NUMBER: US/09/801,876B
 ; CURRENT FILING DATE: 2001-03-09
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 148567
 ; TYPE: DNA
 ; ORGANISM: Human
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)..(148567)
 ; OTHER INFORMATION: n = A,T,C or G
 ; US-09-801-876B-3

Query Match 2.1%; Score 36.6; DB 10; Length 148567;
 Best Local Similarity 46.1%; Pred. No. 80;
 Matches 123; Conservative 0; Mismatches 144; Indels 0; Gaps 0;
 1416 ATATAATCAGAAATTTCCATTTTAAACACACTTCCCATGGTTACATCATCTTAAACCAACAT 1475
 31673 AGATAATCAGTGTGTTTAACTTTTAAATTAAGCAGTAGGAGAAATGACTTTTGGGAACCTT 31732
 1476 TTGCATCTCCTATCTAGCCCAAGTATCTATTGAAAGTGGAACTTGGCCACCTAAATTTAGA 1535
 31733 AGAAATTTGGAAACCTTTTATTTCTATGATTGAATATCAACTATATGTAATTTAGTCTTAAG 31792
 1536 TGATTTTGTGCTGTTTGTGCAAGACACAGTGAAGAAACATGGATAAGACAATTTCTTGT 1595
 31793 GTTATATGCTAGAAACATTTTCAAAAACGAAAGCAGCAGCAATGACATCAAAAATGCAATGT 31852
 1596 CAAAATGAAATATTAATTAATAGATGAATGCACTTGTGAAGCCAGCAGACAGATGAC 1655
 31853 CAAAAGCAATGGTTTAAATAGAAATACATCATTTTAAACAATCTTGAAGTTTAAAGAT 31912
 1656 CCTCAGCTCAACTTTTCCCAATAAAGA 1682
 31913 CCTATAAAATCACAAACCCAGAAGGA 31939

RESULT 13

ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (178)...(2202)
 US-10-162-012-26
 Query Match 2.2%; Score 38.2; DB 9; Length 2326;
 Best Local Similarity 46.1%; Pred. No. 1.4; Mismatches 273; Indels 7; Gaps 3;
 Matches 239; Conservative 0;
 79 TGGAGAACCAAAACAGTGGCGAGAGAGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 138
 303 TGGACTATGCTCCAGAGTGAAGACCAAAAGAGACAGAGTGAAGAGGCTACTTCTGGCTGA 362
 139 CGAGATATGTTGTTTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 197
 363 AGGAACATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 422
 198 GTATATCAATGGCAGACGCTGAAGCAGTATATGTAACAGGTTATGTCCTAGCTGGGCTCA 257
 423 ACATTTCAATGGCTGGCAGGTCAGGTGCTGCTACGGGCTATTTCTGTA---TCAGCTTA 479
 258 GGCACCAATGGATATCTCTTACTGATTTTATGTTGGCTGTTCTTTGCAAAACCTAT 317
 480 TGAACCTTAATGGCTGTTTCTGCTGATGTTGGCTGGCTGATCTCTTACCCATCTACAT 539
 318 CGTTCAAGAGGGTATGACCATGTTAGACCCGTTTACGCAAAATCTATGTAAGAAACGCAAT 377
 540 TGCCTGAGTCCAGGATGACGAGTACCTACGGAAGCGCTTCGGTGGCAGCAAT 599
 378 GGGGGGACTCCTGTTTATCTGCACTGATGGGAGAAATGTTCTGGGCTGCGAGCAATTT 437
 600 CCCCATCATCTGGCTGCTACTCTACTATTTATCTACATCTTACCAAGATCTCGGTAGA 659
 438 CTCGCTTTGGAGCCACCATCAGGTGATCATCG---ATGTTGATATGCAATTTCTGT 494
 660 CATGATGACAGTGGCACTCTTCCAGCAGTCTTCGACCTGGATCTGTACCTGGCCAT 719
 495 CATCATCTCTGCACATCTGTCACATCTGTACACACTGGTGGGAGGCTCTATCTGTGCG 554
 720 AGTTGGGCTACTGGCCATCCTGCTGCTATACACAGTGTGCTGGTGGCTGCTGCTGTGAT 779
 555 CTACACTGATCTGTTCTACCTCTTTTGCATTTTGTAGG 593
 780 CTACAGGATGCGCTGCAGACGCTGATCATGCTTAGG 818

RESULT 11
 S-10-254-869-3
 ; Sequence 3, Application US/10254869
 ; Publication No. US20030027307A1
 ; GENERAL INFORMATION:
 ; APPLICANT: YE, Jane et al
 ; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
 ; FILE REFERENCE: CL001160DIV
 ; CURRENT APPLICATION NUMBER: US/10/254,869
 ; CURRENT FILING DATE: 2002-09-26
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 148567
 ; TYPE: DNA
 ; ORGANISM: Human
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: (1)..(148567)
 ; OTHER INFORMATION: n = A,T,C or G
 ; US-10-254-869-3

Query Match 2.1%; Score 36.6; DB 9; Length 148567;
 Best Local Similarity 46.1%; Pred. No. 80;

Tue Apr 1 13:57:24 2003

us-10-069-541-5.rni

Sequence 18, Appl
Sequence 1, Appl
Sequence 2035, Ap
Sequence 172, App
Sequence 1564, Ap
Sequence 1880, Ap
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 13, Appl
Sequence 12, Appl
Patent No. 5231168
Sequence 14, Appl
Sequence 3, Appl

316 4 US-09-602-877A-18
2101 1 US-08-106-761-1
783 4 US-09-134-001C-2035
11770 4 US-08-961-527-172
933 4 US-09-134-001C-1564
1272 4 US-09-134-001C-1880
3182 4 US-08-971-395-1
3183 1 US-08-413-135-1
4467 1 US-08-565-907A-1
4467 2 US-08-910-551B-1
4467 2 US-08-909-425A-1
35081 2 US-08-752-760A-1
1668 4 US-09-813-872-3
1787 1 US-08-278-630A-13
1980 1 US-08-278-630A-12
3095 6 5231168-1
13011 2 US-08-791-849A-14
65042 4 US-09-784-316-3

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
; US-08-232-463-14

Query Match 2.7%; Score 47.6; DB 1; Length 7218;

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2003, 16:23:55 ; Search time 84 Seconds
(without alignments)
6363.543 Million cell updates/sec

Title: US-10-069-541-5
Perfect score: 1743
Sequence: 1 atggcttcattggaagg.....ctgaagataattacagtga 1743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	47.6	2.7	7218	1 US-08-232-463-14	Sequence 14, Appl
2	41	2.4	2839	4 US-08-595-553A-1	Sequence 1, Appl
3	36.6	2.1	4160	4 US-09-134-218-1	Sequence 1, Appl
4	35.6	2.0	2397	4 US-09-221-017B-272	Sequence 272, App
5	35.2	2.0	2238	1 US-07-841-651-1	Sequence 1, Appl
6	34.8	2.0	1593	4 US-09-134-001C-1673	Sequence 1673, Ap
7	34.6	2.0	25002	4 US-08-961-527-48	Sequence 48, Appl
8	34.2	2.0	2847	4 US-09-484-970B-22	Sequence 22, Appl
9	34.2	2.0	1515	4 US-09-071-035-431	Sequence 431, App
10	34.2	2.0	1803	4 US-09-071-035-429	Sequence 429, App
11	33.8	1.9	3172	1 US-07-741-940-3	Sequence 3, Appl
12	33.8	1.9	3172	1 US-08-289-548A-3	Sequence 3, Appl
13	33.8	1.9	3172	1 US-08-452-654-3	Sequence 3, Appl
14	33.8	1.9	3172	1 US-08-452-658B-3	Sequence 3, Appl
15	33.8	1.9	3172	3 US-08-450-582-3	Sequence 3, Appl
16	33.8	1.9	3172	3 US-08-449-731-3	Sequence 3, Appl
17	33.8	1.9	176373	3 US-09-128-155-17	Sequence 3, Appl
18	33.6	1.9	84495	4 US-09-797-906-3	Sequence 3, Appl
19	33.4	1.9	3593	4 US-09-404-627-1	Sequence 1, Appl
20	33.4	1.9	4205	4 US-08-920-422-17	Sequence 17, Appl
21	33.4	1.9	48974	4 US-09-821-736-1	Sequence 1, Appl
22	33.2	1.9	4396	4 US-08-821-736-1	Sequence 1, Appl
23	32.8	1.9	5935	4 US-08-972-927-2	Sequence 2, Appl
24	32.6	1.9	9936	4 US-08-685-871-1	Sequence 1, Appl
25	32.4	1.9	4739	3 US-08-118-554-18	Sequence 18, Appl
26	32.4	1.9	316	4 US-09-118-627-18	Sequence 18, Appl
27	32.4	1.9	316	4 US-09-118-627-18	Sequence 18, Appl

us-10-069-541-5.rni

Tue Apr 1 13:57:24 2003

```

RESULT 4
US-09-221-017B-272/C
; Sequence 272, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FORSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PPI182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PPI546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PPI2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 272:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2397 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...2397
US-09-221-017B-272

Query Match 2.0%; Score 35.6; DB 4; Length 2397;
Best Local Similarity 67.8%; Pred. No. 0.97; Mismatches 0; Gaps 0;
Matches 50; Conservative 0; Indels 24;

QY 1311 AGGTTATGTTCTGGCCTCTCTCTGAGAACTGAGAGGAGCCATATCTGATCTTCA 1370
Db 1184 AGTTATTTCTGGACTCCCTCTCAATAAATAGGTGAGGAGCCCTATATGCAATCA 1125
QY 1371 GCCCTTGATCTTCT 1384
Db 1124 TATCCTAATCTTCT 1111

```

```

RESULT 5
US-07-841-651-1
; Sequence 1, Application US/07841651
; Patent No. 5410031
; GENERAL INFORMATION:
; APPLICANT: Pajor, Ana M
; TITLE OF INVENTION: Cloning and Functional Expression of a
; Mammalian Na+/Nucleoside Cotransporter: A Member of the
; SGLT Family
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, Ninth Floor
; CITY: Pasadena
; STATE: California
; COUNTRY: USA
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,651
; FILING DATE: 19920224
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandel, Saralynn
; REGISTRATION NUMBER: 31,853
; REFERENCE/DOCKET NUMBER: 8772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 796-4000
; TELEFAX: (818) 795-8321
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2238 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Oryctolagus cuniculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 7..2022
US-07-841-651-1

Query Match 2.0%; Score 35.2; DB 1; Length 2238;
Best Local Similarity 52.8%; Pred. No. 1.2; Indels 0; Gaps 0;
Matches 76; Conservative 0; Mismatches 68;

QY 477 GGATATGCACATTTCTGTCATCATCTCTGCACATCATGCCACTCTGTACACACTGGTGGG 536
Db 522 GAATATTACGCTTCGGTCTATCGCGCTCTGGGCATCACCATGGTTTACACCGTGACAGG 581
QY 537 AGGCTCTATTCTGTGGCCTACACTGATGTCGTTTGCATCTTTTGTAGGCT 596
Db 582 AGGCTGGCAGCGCTGATGTACACAGACACAGTGCAGACCTTTGTCTCATCTCGCGGGGC 641
QY 597 GTGGATCAGCGTCCCTTTGCATT 620
Db 642 CTTTCATCTCACCGGTTACGCCCT 665

RESULT 6
US-09-134-001C-1673
; Sequence 1673, Application US/09134001C
; Patent No. 6380370

```

RESULT 7
S-08-961-527-48/C
Sequence 48, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373

[illegible]

us-10-069-541-5.rni

Tue Apr 1 13:57:24 2003

Db 2151 TTAGATAGAGGACATGTGAACAAATTCATTGAAATTTGATTCACAGATCCATTTC 2210

QY 1671 CACCAATAAGAGCGCTTCT 1691

Db 2211 CAGTGGCAACACAGCAAGCCT 2231

RESULT 9

US-09-071-035-431

; Sequence 431, Application US/09071035

; Patent No. 6448043

; GENERAL INFORMATION:

; APPLICANT: Gil H. Choi

; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 496

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/071,035

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: A. Anders Brookes

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB369P2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 431:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1515 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; US-09-071-035-431

Query Match 2.0%; Score 34.2; DB 4; Length 1515;

Best Local Similarity 54.3%; Pred. No. 2;

Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1506 TGAAGTGGACCTTGGCCACCTAAATAGATGATGCTGCTTTGATGCTGCTTTGCAAGACACAG 1565

Db 1383 TTAAGATAAAATGGCCACCTAAAGTGGATGATGCTGCTTTATTGAAGCAGCCGCGTG 1442

QY 1566 TGAAGAAACATGATGAAGACAATCTTGTCACAAATGAAATATTAATTAATAGATGAAC 1625

Db 1443 GGAAAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1502

QY 1626 TGCACCT 1632

Db 1503 TGGAGTT 1509

RESULT 10

US-09-071-035-429

; Sequence 429, Application US/09071035

; Patent No. 6448043

; GENERAL INFORMATION:

; APPLICANT: Gil H. Choi

; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 429:

SEQUENCE CHARACTERISTICS:

LENGTH: 1803 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-071-035-429

Query Match 2.0%; Score 34.2; DB 4; Length 1803;

Best Local Similarity 54.3%; Pred. No. 2.2;

Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1506 TGAAGTGGACCTTGGCCACCTAAATAGATGATGCTGCTTTGATGCTGCTTTGCAAGACACAG 1565

Db 1643 TTAAGATAAAATGGCCACCTAAAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1702

QY 1566 TGAAGAAACATGATGAAGACAATCTTGTCACAAATGAAATATTAATTAATAGATGAAC 1625

Db 1703 GGAAAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1762

QY 1626 TGCACCT 1632

Db 1763 TGGAGTT 1769

RESULT 11

US-07-741-940-3/c

; Sequence 3, Application US/07741940

; Patent No. 5352775

; GENERAL INFORMATION:

; APPLICANT: ALBERTSEN, HANS

; APPLICANT: ANAND, RAKESH

; APPLICANT: CARLSON, MARY

; APPLICANT: GRODEN, JOANNA

; APPLICANT: HEDGE, PHILIP J.

; APPLICANT: JOSLYN, GEOFF

; APPLICANT: KINZLER, KENNETH

; APPLICANT: MARKHAM, ALEXANDER F.

; APPLICANT: NAKAMURA, YUSUKE

; APPLICANT: THLIVERIS, ANDREW

; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

; NUMBER OF SEQUENCES: 94

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Banner, Birch, McKie & Beckett

; STREET: 1001 G Street, NW

;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20001-4598
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/741,940
;; FILING DATE: 19920109
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kagan, Sarah A.
;; REGISTRATION NUMBER: 32,141
;; REFERENCE/DOCKET NUMBER: 1107.035574
;; TELEPHONE: 202-508-9100
;; TELEFAX: 202-508-9299
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3172 base pairs
;; TYPE: NUCLEIC ACID
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; ORIGINAL SOURCE:
;; ORGANISM: Homo sapiens
;; IMMEDIATE SOURCE:
;; CLONE: DPl(TB2)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..630
;; US-07-741-940-3

Query Match 1.9%; Score 33.8; DB 1; Length 3172;
Best Local Similarity 60.2%; Pred. No. 4.2;
Matches 56; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
Qy 1022 CTGCTGCTGTTATGTCATCAGCAGATCTTCCATCTTGTCCAGCAAGTTCCATGTTGCAC 1081
Db 583 CTTCCTTAGTATGATGCATCTGCATCTCTTTGGACTTGTCTTTAAGGTCCTTGACCACAC 524
Qy 1082 GGACATCTACCAAGTTTCCCTTCAGACAAATG 1114
Db 523 TGTCATCTGGGACTCGTGTCTCAGGAAGAAAG 491

RESULT 12
US-08-289-548A-3/C
; Sequence 3, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allogretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA

;; ZIP: 20001-4598
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/289,548A
;; FILING DATE: 12-AUG-1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kagan, Sarah A.
;; REGISTRATION NUMBER: 32,141
;; REFERENCE/DOCKET NUMBER: 1107.46943
;; TELEPHONE: 202-508-9100
;; TELEFAX: 202-508-9299
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3172 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; ORIGINAL SOURCE:
;; ORGANISM: Homo sapiens
;; IMMEDIATE SOURCE:
;; CLONE: DPl(TB2)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..630
;; US-08-289-548A-3

Query Match 1.9%; Score 33.8; DB 1; Length 3172;
Best Local Similarity 60.2%; Pred. No. 4.2;
Matches 56; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
Qy 1022 CTGCTGCTGTTATGTCATCAGCAGATCTTCCATCTTGTCCAGCAAGTTCCATGTTGCAC 1081
Db 583 CTTCCTTAGTATGATGCATCTGCATCTCTTTGGACTTGTCTTTAAGGTCCTTGACCACAC 524
Qy 1082 GGACATCTACCAAGTTTCCCTTCAGACAAATG 1114
Db 523 TGTCATCTGGGACTCGTGTCTCAGGAAGAAAG 491

RESULT 13
US-08-452-654-3/C
; Sequence 3, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

us-10-069-541-5.rni

Tue Apr 1 13:57:24 2003

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/452,655B
 FILING DATE: 25-MAY-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/289,548
 FILING DATE: 12-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/741,940
 FILING DATE: 08-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A. 32,141
 REGISTRATION NUMBER: 1107.49964
 REFERENCE/DOCKET NUMBER:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3172 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: DPL(TB2)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..630
 US-08-452-655B-3

Query Match 1.9%; Score 33.8; DB 1; Length 3172;
 Best Local Similarity 60.2%; Pred. No. 4.2;
 Matches 56; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1022 CTGCTGCTGTTATGTCATCAGCAGATCTTCCATCTGTCAGCAAGTTCATGTTGCAC 1081
 Db 583 CTTCTTTAGTGGCATCTGCAGTCTCTTTGGACTCTGCTTTAAGGTCCTTGACCAC 524

QY 1082 GGAACATCTACCAGCTTTCCTTCAGACAAATG 1114
 Db 523 TGTCCATCTGGGACTCGTCTTCAGAGAAAG 491

RESULT 14
 US-08-452-655B-3/C
 Sequence 3, Application US/08452655B
 Patent No. 5783666
 GENERAL INFORMATION:
 APPLICANT: ALBERTSEN, HANS
 APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THLIVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 102
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner & Witcoff, Ltd.
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/452,655B
 FILING DATE: 25-MAY-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/289,548
 FILING DATE: 12-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/741,940
 FILING DATE: 08-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A. 32,141
 REGISTRATION NUMBER: 1107.49964
 REFERENCE/DOCKET NUMBER:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3172 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: DPL(TB2)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..630
 US-08-452-655B-3

Query Match 1.9%; Score 33.8; DB 1; Length 3172;
 Best Local Similarity 60.2%; Pred. No. 4.2;
 Matches 56; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1022 CTGCTGCTGTTATGTCATCAGCAGATCTTCCATCTGTCAGCAAGTTCATGTTGCAC 1081
 Db 583 CTTCTTTAGTGGCATCTGCAGTCTCTTTGGACTCTGCTTTAAGGTCCTTGACCAC 524

QY 1082 GGAACATCTACCAGCTTTCCTTCAGACAAATG 1114
 Db 523 TGTCCATCTGGGACTCGTCTTCAGAGAAAG 491

RESULT 15
 US-08-450-582-3/C
 Sequence 3, Application US/08450582
 Patent No. 6114124
 GENERAL INFORMATION:
 APPLICANT: ALBERTSEN, HANS
 APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THLIVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 102
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner & Witcoff, Ltd.
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA

ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,582
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/452,655
FILING DATE: 25-MAY-1995
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3172 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: DP1(TB2)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..530
US-08-450-582-3

Query Match 1.9%; Score 33.8; DB 3; Length 3172;
Best Local Similarity 60.2%; Pred. No. 4.2;
Matches 56; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
Qy 1022 CTGCTGCTGTTATGTCATCAGCAGATCTTCCATCTTGTCAGCAAGTTCCATGTTGCAC 1081
Db 583 CTTCTTAGTATGGCATCTGCAGTCTCTTTGGACTTGTCTTTAAGGTCCTTGACCACAC 524
Db 1082 GGAACATCTACCACTTTCCTTCAGACAAATG 1114
Db 523 TGTCCATCTGGACTCTGCTTCAGGAAGAAAG 491

Search completed: March 31, 2003, 18:30:43
Job time : 146 secs

us-10-069-541-5.rng

Tue Apr 1 13:57:24 2003

Bacillus lichenifo
Human CHOT promote
Drosophila melanog
DNA encoding novel
Human CHOT exon 3
Human foetal liver
Probe #1838 for ge
Human brain expres
Human bone marrow
Probe #1846 for ge
Probe #1931 used t
Probe #1842 used t
Human CHOT exon 4
Human CHOT exon 5
Human CHOT exon 2
Human foetal liver
Probe #11735 for g
Human brain expres
Human bone marrow
Probe #11054 for g
Probe #15074 used t
Probe #6832 used t
Human spliced tran
Mouse spliced tran
Gene encoding a su
Pyrococcus abyssi
Arabidopsis thalia
Human sbg1007026SG
Human transport pr
Rat NIS nucleotide
Rat sodium/Iodide
Yeast AOD9604-asso
Novel human transp
Novel human transp
Human sodium-sugar
Candida albicans p

10	242.6	13.9	1461	24	ABK73210
11	180.8	10.4	10140	22	AAH49201
12	179.6	10.3	4223	23	ABL29568
13	167.6	9.6	1094	23	AAS82193
14	163	9.4	240	22	AAH49203
15	155	8.9	455	22	ABA53620
16	155	8.9	455	22	ABA23372
17	155	8.9	455	22	AAK01883
18	155	8.9	455	22	AAK27341
19	155	8.9	455	22	AAI11913
20	155	8.9	455	22	AAI33245
21	155	8.9	455	22	AAI01851
22	154.6	8.9	300	22	AAH49204
23	150.8	8.7	240	22	AAH49205
24	108	6.2	180	22	AAH49202
25	72	4.1	96	22	ABA66205
26	72	4.1	96	22	ABA33269
27	72	4.1	96	22	AAK14621
28	72	4.1	96	22	AAK40366
29	72	4.1	96	22	AAI21121
30	72	4.1	96	22	AAI06841
31	72	4.1	96	22	ABN37501
32	60	3.4	60	24	ABN54094
33	53.8	3.1	10732	21	ABN10594
34	51.8	3.0	34980	22	AAF86431
35	42.4	2.4	11220	21	AAC48974
36	41.6	2.4	2028	24	ABL57735
37	41.2	2.4	2200	22	AAF27713
38	41.2	2.4	1857	22	AAF79537
39	41	2.4	1857	22	AAI88971
40	41	2.4	2839	18	AAI88971
41	39.8	2.3	4590	22	AAH24065
42	39.6	2.3	2028	22	AAF84022
43	39.6	2.3	2456	22	AAF84023
44	38.2	2.2	2326	24	ABK10764
45	38.2	2.2	6426	21	AAA57920

ALIGNMENTS

RESULT 1
AAF81712
ID AAF81712 standard; cDNA; 1743 BP.
XX AAF81712;
XX 01-JUN-2001 (first entry)
XX Human high affinity choline transporter protein encoding cDNA.
XX High affinity choline transporter; cho-1; Alzheimer's disease;
XX diagnosis; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 1..1743
XX /*tag= a
XX /product= "high affinity choline transporter"
XX WO200116315-A1.
XX 08-MAR-2001.
XX 18-AUG-2000; 2000WO-JP05545.
XX 27-AUG-1999; 99JP-0240642.
XX 27-DEC-1999; 99JP-0368991.
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX Haga T, Okuda T;
PI

OM nucleic - nucleic search, using sw model
Run on: March 31, 2003, 15:07:07 ; Search time 396 Seconds
(without alignments)
9912.205 Million cell updates/sec

Title: US-10-069-541-5
Perfect score: 1743
Sequence: 1 atggtttccatgtgaagg.....ctgaagataattacagtga 1743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2185239 seqs, 112599159 residues 4370478
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	N_Geneseq_101002.*
1:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1743	100.0	1743	22	AAF81712 Human high affinity
2	1743	100.0	1743	22	AAH49207 Human CHOT encodin
3	1394.2	80.0	1743	22	AAF81711 Rat high affinity
4	1373.4	78.8	1743	22	AAF81713 Mouse high affinity
5	1373.4	78.8	4938	22	AAD02457 Mouse P4P6B1 OMA (
6	630.8	36.2	8760	22	AAH49206 Human CHOT exons 6
7	363.8	20.9	1731	22	AAF81710 C. elegans high af
8	279.6	16.0	386	22	AAD02461 Mouse P4P6B1 cDNA
9	266.6	15.3	1729	23	ABL29569 Drosophila melanog

XX WPI; 2001-226688/23.
 DR P-PSDB; AAB74665.
 XX
 PT New rat and human spinal cord high affinity choline transporters,
 PT useful in diagnosis of Alzheimer's disease and screening promoters as
 XX drugs for treating Alzheimer's disease
 XX
 PS Claim 9; Page 71-75; 90pp; Japanese.
 XX
 CC The present sequence encodes a human (Homo sapiens) high affinity
 CC choline transporter protein designated cho-1. The cho-1 protein has
 CC nootropic and neuroprotective activities. The cho-1 polynucleotide and
 CC protein can be used for the diagnosis of diseases related to the
 CC expression of cho-1 by comparing the cho-1 polynucleotide sequence in a
 CC sample to that of a control. Drug compositions containing the cho-1
 CC protein or expression promoters or inhibitors of cho-1 are useful for
 CC treating disorders characterised by abnormal levels of cho-1, such as
 CC Alzheimer's disease.
 XX
 CC
 CC Sequence 1743 BP; 412 A; 393 C; 406 G; 532 T; 0 other;
 CC
 CC Query Match 100.0%; Score 1743; DB 22; Length 1743;
 CC Best Local Similarity 100.0%; Pred. No. 0;
 CC Matches 1743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC
 CC 1 ATGGCTTTCCATGTGGAAGGACTAGTACTATCATCGTGTCTTACCTTCTAATTTTGGTG 60
 CC |||||
 CC 1 ATGGCTTTCCATGTGGAAGGACTAGTACTATCATCGTGTCTTACCTTCTAATTTTGGTG 60
 CC |||||
 CC 61 GTTGGAAATATGGCTGCCCTGGAGAACCAAAAACAGTGGCAGCGCAGAGCGCAGCGAA 120
 CC |||||
 CC 61 GTTGGAAATATGGCTGCCCTGGAGAACCAAAAACAGTGGCAGCGCAGAGCGCAGCGAA 120
 CC |||||
 CC 121 GCCATCATAGTTGGTGGCGGAGATATGGTTATTGGTGGTGGATTTACCATGACAGCT 180
 CC |||||
 CC 121 GCCATCATAGTTGGTGGCGGAGATATGGTTATTGGTGGTGGATTTACCATGACAGCT 180
 CC |||||
 CC 181 ACCTGGTGGAGAGGATATATCAATGCGACAGCTGAAGCAGTTTATGACCGAGTTAT 240
 CC |||||
 CC 181 ACCTGGTGGAGAGGATATATCAATGCGACAGCTGAAGCAGTTTATGACCGAGTTAT 240
 CC |||||
 CC 241 GGCCTAGCTTGGCTCAGGCACCAATTTGATATCTTCTAGTCTGATTTAGTGGCCTG 300
 CC |||||
 CC 241 GGCCTAGCTTGGCTCAGGCACCAATTTGATATCTTCTAGTCTGATTTAGTGGCCTG 300
 CC |||||
 CC 301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGACCATGTTAGACCGCTTTTCAGCAA 360
 CC |||||
 CC 301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGACCATGTTAGACCGCTTTTCAGCAA 360
 CC |||||
 CC 361 ATCTATGGAACACGATGGGGGACTCTGTTTATCTCTGACATGATGGGAGAAATGTC 420
 CC |||||
 CC 361 ATCTATGGAACACGATGGGGGACTCTGTTTATCTCTGACATGATGGGAGAAATGTC 420
 CC |||||
 CC 421 TGGGCTGACGCAATTTTCTCTGCTTTGGGAGCCACCATGACGCTGATCATCGATGGAT 480
 CC |||||
 CC 421 TGGGCTGACGCAATTTTCTCTGCTTTGGGAGCCACCATGACGCTGATCATCGATGGAT 480
 CC |||||
 CC 481 ATGCACATTTCTGTCATCATCTCTGACATCAATTTGCCACTCTGTACACATGTTGGGAGG 540
 CC |||||
 CC 481 ATGCACATTTCTGTCATCATCTCTGACATCAATTTGCCACTCTGTACACATGTTGGGAGG 540
 CC |||||
 CC 541 CTCTATTCTGTGGCCCTACACTGATGTCGTTTTCAGCTCTTTTGGCAATTTTGTAGGGCTGTG 600
 CC |||||
 CC 541 CTCTATTCTGTGGCCCTACACTGATGTCGTTTTCAGCTCTTTTGGCAATTTTGTAGGGCTGTG 600
 CC |||||
 CC 601 ATCAGCGTCCCTTTTGCATTTGTACATCTCTGAGTGGCAGACATCGGGTTTCACTGCTGTG 660
 CC |||||
 CC 601 ATCAGCGTCCCTTTTGCATTTGTACATCTCTGAGTGGCAGACATCGGGTTTCACTGCTGTG 660
 CC |||||
 CC 661 CATGCCAATACCAAAAGCCGTGGTGGAACTGTGTGACTATCTGAACTCTACTCTTGG 720
 CC |||||
 CC 661 CATGCCAATACCAAAAGCCGTGGTGGAACTGTGTGACTATCTGAACTCTACTCTTGG 720
 CC |||||

QY 721 CTGTGATAGTCTTCTGTTGATGCTGGTGGAAATCCCATGGCAAGCATACTTTTCAGAGG 780
 DB |||||
 QY 721 CTGTGATAGTCTTCTGTTGATGCTGGTGGAAATCCCATGGCAAGCATACTTTTCAGAGG 780
 DB |||||
 QY 781 GTTCTCTCTTCTCTCAGCCACCTATGCTCAAGGCTGTCTCTTCTCTGCGAGCTTTTCGGG 840
 DB |||||
 QY 781 GTTCTCTCTTCTCTCAGCCACCTATGCTCAAGGCTGTCTCTTCTCTGCGAGCTTTTCGGG 840
 DB |||||
 QY 841 TGCCTGGTGTATGGCCATCCCGAGCCATACTCATTTGGGGCCATTGGAGCATCAACAGACTGG 900
 DB |||||
 QY 841 TGCCTGGTGTATGGCCATCCCGAGCCATACTCATTTGGGGCCATTGGAGCATCAACAGACTGG 900
 DB |||||
 QY 901 AACCCAGACTGATATGGCTTCCAGATCCCAAGACTACAGAGAGGAGGAGCATGATTTTA 960
 DB |||||
 QY 901 AACCCAGACTGATATGGCTTCCAGATCCCAAGACTACAGAGAGGAGGAGCATGATTTTA 960
 DB |||||
 QY 961 CCAATTTCTTGCAGTATCTGCTGCTGTGTATATTTCTTTTCTTGGTCTTGGTGCAGTT 1020
 DB |||||
 QY 961 CCAATTTCTTGCAGTATCTGCTGCTGTGTATATTTCTTTTCTTGGTCTTGGTGCAGTT 1020
 DB |||||
 QY 1021 TCTGCTGCTTATGTCATCAGCAGATTTCTTCCATCTTGTCCAGCAAGTTCATGTTTGA 1080
 DB |||||
 QY 1021 TCTGCTGCTTATGTCATCAGCAGATTTCTTCCATCTTGTCCAGCAAGTTCATGTTTGA 1080
 DB |||||
 QY 1081 CGGAAATCTACAGCTTTCTTCTTCCAGCAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
 DB |||||
 QY 1081 CGGAAATCTACAGCTTTCTTCTTCCAGCAAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
 DB |||||
 QY 1141 ATGCGAATCACAGTGTGTTGTTGGAGCATCTGCAACAGCCATGGCTTGTGAGGAA 1200
 DB |||||
 QY 1141 ATGCGAATCACAGTGTGTTGTTGGAGCATCTGCAACAGCCATGGCTTGTGAGGAA 1200
 DB |||||
 QY 1201 ACTGTGTATGGGCTGTGATACCTGACCTTGTGTTTACATCTTATCTTCCCTCCAG 1260
 DB |||||
 QY 1201 ACTGTGTATGGGCTGTGATACCTGACCTTGTGTTTACATCTTATCTTCCCTCCAG 1260
 DB |||||
 QY 1261 CTGCTTTGTGTAATCTTCTTAAAGGAAACCAACCTATGGGGCCGTGCGAGTTATGTT 1320
 DB |||||
 QY 1261 CTGCTTTGTGTAATCTTCTTAAAGGAAACCAACCTATGGGGCCGTGCGAGTTATGTT 1320
 DB |||||
 QY 1321 TCTGGCTCTTCTCTGAGAAATAACTGGAGGGAGCCATATCTATCTTCCAGCCCTGATC 1380
 DB |||||
 QY 1321 TCTGGCTCTTCTCTGAGAAATAACTGGAGGGAGCCATATCTATCTTCCAGCCCTGATC 1380
 DB |||||
 QY 1381 TTCTACCTCTGCTATTTACCTCTGATGATATGATATATATATCAGAAATTTCCATTTAA 1440
 DB |||||
 QY 1381 TTCTACCTCTGCTATTTACCTCTGATGATATGATATATATATCAGAAATTTCCATTTAA 1440
 DB |||||
 QY 1441 ACATTTGCCATGGTTACATCAATCTTAAACCAACATTTTGCATCTCTATCTAGCCCAAGTAT 1500
 DB |||||
 QY 1441 ACATTTGCCATGGTTACATCAATCTTAAACCAACATTTTGCATCTCTATCTAGCCCAAGTAT 1500
 DB |||||
 QY 1501 CTATTTGAAAGTGGAACTTTCCCACTTAAATAGATGATTTTGTGATGTTGTTGCAAGA 1560
 DB |||||
 QY 1501 CTATTTGAAAGTGGAACTTTCCCACTTAAATAGATGATTTTGTGATGTTGTTGCAAGA 1560
 DB |||||
 QY 1561 CACAGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
 DB |||||
 QY 1561 CACAGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
 DB |||||
 QY 1621 GAATTTGCACTTTGTGAGGCGCAGCAGAGCATGACCTCTAGCTCAACTTTTCCCAATAAA 1680
 DB |||||
 QY 1621 GAATTTGCACTTTGTGAGGCGCAGCAGAGCATGACCTCTAGCTCAACTTTTCCCAATAAA 1680
 DB |||||
 QY 1681 GAGGCTTCTCTTGTGATCTTGTATCCAGTCCAGAGGCTGTGGGACTGAAGATAATTTACAG 1740
 DB |||||
 QY 1681 GAGGCTTCTCTTGTGATCTTGTATCCAGTCCAGAGGCTGTGGGACTGAAGATAATTTACAG 1740
 DB |||||
 QY 1741 TGA 1743
 DB |||||
 QY 1741 TGA 1743
 DB |||||

Tue Apr 1 13:57:24 2003

RESULT 2
ID AAH49207
XX AAH49207 standard; cDNA; 1743 BP.
AC AAH49207;
XX
DT 26-NOV-2001 (first entry)
XX
DE Human CHOT encoding cDNA.
XX
KW CHOT; human; choline transporter; chromosome 2q11-13; nootropic;
KW neuroprotective; gene therapy; antisense therapy; degenerative disease;
KW cognitive disorder; Alzheimer's disease; ss.
XX
XX Homo sapiens.
OS
XX DE10009055-Al.
XX
XX 30-AUG-2001.
XX
XX 28-FEB-2000; 2000DE-1009055.
XX
XX 28-FEB-2000; 2000DE-1009055.
XX
XX (BRUE/) BRUESS M.
XX (BOEN/) BOENISCH H.
XX
XX Bruess M, Boenisch H;
PI
XX WPI; 2001-590709/67.
DR P-PSDB: AAB86837.
XX
XX A new gene encoding human choline transporter, designated hCHOT is
XX located on chromosome 2q11-13 and is useful to treat degenerative
XX disorders such as Alzheimer's disease
XX
XX Disclosure; Page 11; 12pp; German.
XX
XX This invention describes a novel gene encoding human choline transporter,
XX designated hCHOT which is located on chromosome 2q11-13. The products of
XX the invention have neurotropic and neuroprotective activity and can be used
XX for gene or antisense therapy. (i) is used to treat degenerative disease,
XX particularly cognitive disorders such as Alzheimer's disease. Sense and
XX antisense oligonucleotides derived from the gene may be used in
XX diagnostics and other techniques. This sequence encodes the human CHOT
XX protein described in the invention.
XX
XX Sequence 1743 BP; 412 A; 393 C; 406 G; 532 T; 0 other;
XX
XX
XX Query Match 100.0%; Score 1743; DB 22; Length 1743;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ATGGCTTCCATGGAAGGACTGATAGCTATCATCGTGTCTACCTTCTAAATTTGCTG 60
XX
XX 1 ATGGCTTCCATGGAAGGACTGATAGCTATCATCGTGTCTACCTTCTAAATTTGCTG 60
XX
XX 61 GTTGAATATGGCTGCTGGAGAACCAAAACAGTGGCAGCGCAGAGCGAGCGAA 120
XX
XX 61 GTTGAATATGGCTGCTGGAGAACCAAAACAGTGGCAGCGCAGAGCGAGCGAA 120
XX
XX 121 GCCATCATAGTTGGTGGCCGAGATATGTTTATTTGTTGGTGGATTACCATGACAGCT 180
XX
XX 121 GCCATCATAGTTGGTGGCCGAGATATGTTTATTTGTTGGTGGATTACCATGACAGCT 180
XX
XX 181 ACCTGGCTGGAGGAGGTATATCAATGGCAGCTGAGCAGTTTATGTACCAAGTTAT 240
XX
XX 181 ACCTGGCTGGAGGAGGTATATCAATGGCAGCTGAGCAGTTTATGTACCAAGTTAT 240
XX
XX 241 GGCTAGCTGGCTCAGGACCAATTTGGATATCTCTTAGTCTGATTTAGTGCGCTG 300
XX
XX 241 GGCTAGCTGGCTCAGGACCAATTTGGATATCTCTTAGTCTGATTTAGTGCGCTG 300
XX

301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTTCAGCAA 360
301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTTCAGCAA 360
361 ATCTATGGAAGACGATGGGCGGACTCCTCTTTATTTCTGCTGCTGATGGGAGAAATGTTTC 420
361 ATCTATGGAAGACGATGGGCGGACTCCTCTTTATTTCTGCTGCTGATGGGAGAAATGTTTC 420
421 TGGGCTGACGCAATTTTCTCTGCTTGGGAGCCACCATGAGCTGATCATGATGTTGGAT 480
421 TGGGCTGACGCAATTTTCTCTGCTTGGGAGCCACCATGAGCTGATCATGATGTTGGAT 480
481 ATGACATTTCTGTCATCATCTGTCATCTATTTGCCACTCTGTACACACTGGTGGAGGG 540
481 ATGACATTTCTGTCATCATCTGTCATCTATTTGCCACTCTGTACACACTGGTGGAGGG 540
541 CTCTATTCTGTGGCCCTACACTGATGCTGCTCAGCTCTTTTGCATTTTGTAGGCTGTGG 600
541 CTCTATTCTGTGGCCCTACACTGATGCTGCTCAGCTCTTTTGCATTTTGTAGGCTGTGG 600
601 ATCAGGCTCCCTTTTGCATTTGTCATCTCTGAGTGGCAGACATCGGTTTCACTGCTGTG 660
601 ATCAGGCTCCCTTTTGCATTTGTCATCTCTGAGTGGCAGACATCGGTTTCACTGCTGTG 660
661 CATGCCAAATAGCAAAAGCCGTGGGTGGAACTGTTGACTCATCTGAACTTACTTCTGG 720
661 CATGCCAAATAGCAAAAGCCGTGGGTGGAACTGTTGACTCATCTGAACTTACTTCTGG 720
721 CTTGATAGTTTCTCTCCTCAGCACCTATGCTCAAGTGGTCTCTTCTGCGAGCTTTCGGG 840
721 CTTGATAGTTTCTCTCCTCAGCACCTATGCTCAAGTGGTCTCTTCTGCGAGCTTTCGGG 840
841 TGCCTGGTGGCCATCCCGACCATCTACTTGGGCGCCATTTGGAGCATCAACAGACTGG 900
841 TGCCTGGTGGCCATCCCGACCATCTACTTGGGCGCCATTTGGAGCATCAACAGACTGG 900
901 AACGAGACTGATATGGGCTTCCAGATCCCAAGACTACAGAGAGGAGGAGCATGATTTTA 960
901 AACGAGACTGATATGGGCTTCCAGATCCCAAGACTACAGAGAGGAGGAGCATGATTTTA 960
961 CCAATTTCTGCGATGATCTCTGCGCTGCTGATATTTCTTCTTGGTGGTGGTGGT 1020
961 CCAATTTCTGCGATGATCTCTGCGCTGCTGATATTTCTTCTTGGTGGTGGTGGT 1020
1021 TCTGCTGCTGTTATGTCATGACGAGATTTCTCCATCTTGTGAGCAAGTTTCCATGTTTGA 1080
1021 TCTGCTGCTGTTATGTCATGACGAGATTTCTCCATCTTGTGAGCAAGTTTCCATGTTTGA 1080
1081 CGGACATCTACAGCTTCTCTTCCAGCAAAATGCTTGGGCAAAAGAAATGTTTGGGTT 1140
1081 CGGACATCTACAGCTTCTCTTCCAGCAAAATGCTTGGGCAAAAGAAATGTTTGGGTT 1140
1141 ATGCGAATCACAGTGTGTTGTTGGAGCATCTGCAACAGCCATGCGCTTGTGAGCAAA 1200
1141 ATGCGAATCACAGTGTGTTGTTGGAGCATCTGCAACAGCCATGCGCTTGTGAGCAAA 1200
1201 ACTGTTGATGGGCTGTTGCTGCTGAGCTTCTGAGCTTGTGATCTTATCTTCCCGAG 1260
1201 ACTGTTGATGGGCTGTTGCTGCTGAGCTTCTGAGCTTGTGATCTTATCTTCCCGAG 1260
1261 CTGCTTTGTGCTACTCTTTGTTAAAGGGAACCAACACTTATGGGCGCGTGTATGTT 1320
1261 CTGCTTTGTGCTACTCTTTGTTAAAGGGAACCAACACTTATGGGCGCGTGTATGTT 1320
1321 TCTGCGCTCTTCTCTGAGAAATTAACCTGGAGGAGGAGCCATATCTGATCTTTCAGCCCTTGATC 1380
1321 TCTGCGCTCTTCTCTGAGAAATTAACCTGGAGGAGGAGCCATATCTGATCTTTCAGCCCTTGATC 1380

KW	diagnosis; ss.	
XX	Mus musculus.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	1..1743
FT		/*tag= a
XX		/product= "high affinity choline transporter"
PN	W0200116315-Al.	
XX		
PD	08-MAR-2001.	
XX		
PF	18-AUG-2000; 2000WO-JP05545.	
XX		
PR	27-AUG-1999; 99JP-0240642.	
PR	27-DEC-1999; 99JP-0368991.	
XX	(NISC-) JAPAN SCI & TECHNOLOGY CORP.	
PA	Haga T, Okuda T;	
PI	WPI; 2001-226688/23.	
XX	P-PSDB; AAB74666.	
DR	New rat and human spinal cord high affinity choline transporters,	
XX	useful in diagnosis of Alzheimer's disease and screening promoters as	
PT	drugs for treating Alzheimer's disease	
PT	Claim 12; Page 78-82; 90pp; Japanese.	
XX		
PS	The present sequence encodes a mouse (Mus musculus) high affinity	
XX	choline transporter protein designated cho-1. The cho-1 protein has	
CC	neurotropic and neuroprotective activities. The cho-1 polynucleotide and	
CC	protein can be used for the diagnosis of diseases related to the	
CC	expression of cho-1 by comparing the cho-1 polynucleotide sequence in a	
CC	sample to that of a control. Drug compositions containing the cho-1	
CC	protein or expression promoters or inhibitors of cho-1 are useful for	
CC	treating disorders characterised by abnormal levels of cho-1, such as	
CC	Alzheimer's disease.	
XX		
SQ	Sequence 1743 BP; 407 A; 410 C; 409 G; 517 T; 0 other;	
	Query Match 78.8%; Score 1373.4; DB 22; Length 1743;	
	Best Local Similarity 86.7%; Pred. No. 0;	
	Matches 1512; Conservative 0; Mismatches 231; Indels 0; Gaps 0;	
QY	1 ATGCTTTCCATGTGGAGGACTGATAGCTATCATCGTGTGTTCTACCTTCTTAATTTGCTG 60	
DB	1 ATGCTTTCCACGTAGAAGGACTGTAGCTATTATCTCTTCTTCTTATTTCTG 60	
QY	61 GTTGAATATGGCTGCTCGCTGGAGAACCAAAACAGTGGCAGCGAGAGCGAGCGAA 120	
DB	61 GTTGAATATGGCTGCTCGATGAAACCAAAACAGCGGCAACCCAGAGACGACAGTCAA 120	
QY	121 GCCATCATAGTTGGTGGCCGAGATATGTTTATTTGGTGGTGGTATTTACCATGACAGT 180	
DB	121 GCCATCATAGTTGGGCGCGGTGACATGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGT 180	
QY	181 ACCTGGGTGGAGGAGGAGGTATATCAATGGCAGACAGTGAAGCAGTTTATGTACCAAGTAT 240	
DB	181 ACCTGGGTGGAGGAGGAGGTATATCAATGGCAGACAGTGAAGCAGTTTATGTACCAAGTAT 240	
QY	241 GGCCTAGCTTGGGCTCAGGCACCAATTTGATATTTCTTAGTCTGATTTAGTGGCCCTG 300	
DB	241 GGCCTAGCTTGGGCTCAGGCACCAATTTGATATTTCTTAGTCTGATTTAGTGGCCCTG 300	
QY	301 TTCTTTTGAACCACTATGCTTCAAGGGGTATGTACCATGTTAGACCCGTTTCAGCAA 360	
DB	301 TTTTGTGCAACCTATGCTTCAAGGGGTATGTACCATGTTAGACCCGTTTCAGCAA 360	
QY	361 ATCTATGGAACAGCATGGCGGACTCTCTGTTTATCTCTGCACTGATGGAGAAATGTTTC 420	
DB	361 ATCTATGGAACAGCATGGCGGACTCTCTGTTTATCTCTGCACTGATGGAGAAATGTTTC 420	

RESULT 4
AAF81713
ID AAF81713 standard; cDNA; 1743 BP.
XX
AC AAF81713;
XX
DT 01-JUN-2001 (first entry)
XX
DE Mouse high affinity choline transporter protein encoding cDNA.
XX
KW High affinity choline transporter; cho-1; Alzheimer's disease;

Db	361	ATCTATGGAAGCGCATGGGTGGCTGTCTTCTATCCCTGCACCTGATGGAGAGATGTTCT	420
Qy	421	TGGGCTGCAGCAAAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCATGTGGAT	480
Db	421	TGGGCTGCAGCAAAATTTTCTCTGCAATAGGGCCACCATCAGCGTGATCATGTGGAT	480
Qy	481	ATGCACATTTCTGTCATCATCTCTGCACTCATTTGCACTCTGTACACACTGGTGGGAGGG	540
Db	481	GTGAACATATCCGGTCATTTGCTCTGCACTCATTTGCCATTCCTTATACCCTAGTGGGTGG	540
Qy	541	CTCTATTCTGTGGGCTACACATGATGCTGCTTCACTCTTTTGCATTTTGTAGGCGTGTGG	600
Db	541	CTCTACTCTGTGGCATATCATGATTTGTCCAGCTATTCTGCATTTTATAGGACTGTGG	600
Qy	601	ATCAGCGTCCCTTTTGGCATTTGTCATCTCAGCTCTGTCAGTCGCAGACATCGGGTTCACTGCTGTG	660
Db	601	ATCAGTGTCCCTTTTGGCCTGTGCATCTCTCAGTCAACGACATCGGATTCACAGCTGTG	660
Qy	661	CATGCCAAATACCAAAAGCCGTGCTGGGAACCTGTTGACTCATCTGAAGTCTACTCTTTGG	720
Db	661	CATGCTAAATACCAAGATCCCTGGCTGGGAACCATTAATCAGTTGAAGTCTACACCTGG	720
Qy	721	CTTGATAGTTTCTGTTGTTGATGCTGGGTGGAAATCCCATGGCAAGCATACTTTTCAGAG	780
Db	721	CTTGATAAATTTCTGTTATTGATGCTGGGTGGAAATCCCATGGCAAGCCTACTTCCAGAG	780
Qy	781	GTCTCTCTTCTTCCCTCAGCCACTATGCTCAAGTGTGTCCTTCTCTGGCAGCTTTCGGG	840
Db	781	GTCTCTCTTCTTCCCTCAGCCACTATGCTCAAGTGTGTCCTTCTCTGGCAGCTTTCGGG	840
Qy	841	TGCTGTGGTATGGCCATCCAGCCATACTCATTTGGGGCCATTTGGAGCATCAACAGACTGG	900
Db	841	TGCTGTGGTATGGCTCTACCCGCCATATGCATAGGAGCTATTGGAGCTTCCACAGACTGG	900
Qy	901	AACCACACTGCATATGGGCTTCCAGATCCCAAGACTACAGAAGGCGAGACATGATTTTA	960
Db	901	AACCACACTGCCTTACGGGTATCCAGATCCCAAGACTAAAGGAGGAAGCAGACATGATTC	960
Qy	961	CCAATTTGTTCTGCAGTATCTCTGCCCTGTGTATTTTCTTTTCTGCTGTGGGCGAGTT	1020
Db	961	CGGATGTTTCTGCAGTACCTTGCCCTGTGTACATCTCTTCTTGGGCTTGGGTGCTGTT	1020
Qy	1021	TCGTGCTGCTGTTATGTCATCAGCAGATTTCTCCATCTGTGTGTCAGCAAGTTTCCATGTTGA	1080
Db	1021	TCAGTGTGCTGATGTCCTCAGCTGACTCGTCCATCTCTGTGGCGAGTTCTATGTTTGT	1080
Qy	1081	CGGAACATCTACACGCTTTCCCTTCAGACAAAATGCTTCGACAAAAGAAATCTTTGGGTT	1140
Db	1081	CGGAATATCTACACGCTTTCCCTTCAGACAAAATGCTACAGACAGGAAATGTGTGGGTC	1140
Qy	1141	ATGCCAAATCACAGTGTGTTGTTGTGGAGCATCTGCAACAGCCATGCCCCCTGTGACGAAA	1200
Db	1141	ATGAGGATCATCTGCTGTGTGTCTGGAGCATCTGCAACAGCCATGCTTTGTGTGACGAAG	1200
Qy	1201	ACTGTGTATGGGCTCTGCTGACTCAGTTCTGACCTTGTTTACATCGTTATCTTCCCCAG	1260
Db	1201	ACTGTGTATGGGCTCTGCTGACTCAGTTCTGACCTTGTTTACATCATCATCTTCCACAG	1260
Qy	1261	CTGCTTTGTGTACTCTTTGTTTAAAGGAACCAACACTATGGGGCGTGGCAGGTTATGTT	1320
Db	1261	CTGCTCTGTGTACTCTTTCATCAAAGGAACCAACACTTATGGGCAAGTTTCTGTGTTATAT	1320
Qy	1321	TCTGGCCTCTTCTTGAGAAATACCTGGAGGAGCCATATCTGTATCTTCAGCCCTTGATC	1380
Db	1321	TTTGGACTATTTCTTGAGAAATACCTGGAGGAGCCATATCTATCTTCAGCCCTTAATC	1380
Qy	1381	TTCTACCCCTGGCTATATACCCCTGATGATAATGGTATATATAATACGAAATTTCCATTTAAA	1440
Db	1381	TTCTACCCCTGGTATATCTCTGACAAGAAATGGTATATACATACAGAGTTTCCCATTTAAA	1440
Qy	1441	ACACTTCCCATGGTTACATCATCTTACCAACAAATTTGCATCTCCTATCTAGCCAAGTAT	1500
Db	1441	ACTCTCTCCATGGTTACTCATCTTACCAACAAATTTGTTTCTTATCTAGCCAAGTAT	1500

Qy	1501	CTATTGAAAGTGGAAACCTTGCCACCTAAATTAGATGTATTTCATGCTGTGTGTCGAAGA	1560
Db	1501	CTATTGAAAGTGGAAACCTTGCCCTCCAAAATTAGATGTATTTCATGCTGTGTGCGCAAGG	1560
Qy	1561	CACAGTGAAGAAACATGGATAAGACAAATCTTGTCAAAAATTGAAAATTTAAATTAGAT	1620
Db	1561	CACAGTGAAGAAACATGGACAAGACCAATCTAGTCAGAAATGAAAATATCAAAATTAAT	1620
Qy	1621	GAACTTGCACCTTGTGAAGCCAGCAGACGACATGACCCCTCAGCTCAACTTTTCACCAATAAA	1680
Db	1621	GAACTTGCACCTGTGAAACCTCGGCAGAGCCTAACCCCTCAGTTCAACTTTTCACCAATAAG	1680
Qy	1681	GAGCCCTCCCTTGATGTTGATTCAGTCCAGNAGGGCTCGGGACCTGAAGATAATTTACAG	1740
Db	1681	GAGCCCTCCCTTGATGTTGATTCAGTCCGCGGGGGTCTGGGACCTGAAGATAACTTACAA	1740
Qy	1741	TGA	1743
Db	1741	TGA	1743

RESULT 5
AAD02457

ID AAD02457 standard; cDNA; 4938 bp.

AA
AC
AAD02457;XX
DT 24-APR-20XX
DE MOISE DAD

XX

KW
mouse, OM
fuel metalXX
KW anorectic

OS Mus sp.
XX

FH	Key
FT	CS

ET

misc_feat

THE

XX
PN WO200007895XX
28-DEC-200

XX

XX
007-NOV 67XX
22-JUN-1995

PA (AMYL-) AM
XX

Sierzega M

WPI; 2001 -

XX
P-PSUB; AANovel obes
treatment

—

Claim 2; F

The presen

mouse adipose
fingerprin

OMA is used in metabolism

the transc.

1

Tue Apr 1 13:57:24 2003

PT New rat and human spinal cord high affinity choline transporters, as
 PT useful in diagnosis of Alzheimer's disease and screening promoters as
 PT drugs for treating Alzheimer's disease -

XX Claim 3; Page 57-62; 90pp; Japanese.

XX The present sequence encodes a Caenorhabditis elegans high affinity
 CC choline transporter protein designated cho-1. The cho-1 protein has
 CC nontropic and neuroprotective activities. The cho-1 polynucleotide
 CC protein can be used for the diagnosis of diseases related to the
 CC expression of cho-1 by comparing the cho-1 polynucleotide sequence in a
 CC sample to that of a control. Drug compositions containing the cho-1
 CC protein or expression promoters or inhibitors of cho-1 are useful for
 CC treating disorders characterised by abnormal levels of cho-1, such as
 CC Alzheimer's disease.

XX Sequence 1731 BP; 428 A; 373 C; 427 G; 503 T; 0 other;

Query Match 20.9%; Score 363.8; DB 22; Length 1731;
 Best Local Similarity 55.1%; Pred. No. 1.4e-95;
 Matches 862; Conservative 0; Mismatches 637; Indels 66; Gaps 5;

QY 19 GGACTGATAGTATCATCTGCTGCTTACCTCTTAATTTGCTGTTGGAATATGGCTGCC 78
 DB 16 GGTATCGTGGCCATGTTGCTTCTAGCTGCTCATCTTCTGTTGGAATATGGCTGCC 75
 QY 79 TGGAGAACAAAA-----ACAGTGGCAGCGCAGAGAGCGCAGCGAGCCATC 126
 DB 76 AGAAATCGAAAGATTCAAAGAGCTTGAATCAGAGCGCGCGGACGAGAGGTG 135
 QY 127 ATAGTTGGTGGCGGAGATTTGTTTATTTGTTGGTGGATTTACATCAGCTACCTGG 186
 DB 136 ATGTTAGCTGGAGAACATCGAACTCTTCTCGGAATTTTCAATGACTGCCACGTGG 195
 QY 187 GTCGAGGAGGTATATCAATGGCAGCTGAAGCAGTTTATGTACCAAGTTTATGGCCTA 246
 DB 196 GTTGGCGGTGCTTATATCAATGGAACCGCGGAGGCTCTGTATATGGAGT-----CTC 249
 QY 247 GCTTGGCTCAGGCAACCAATTTGATTTCTTATGTCGTGATTTTATGGTGGCTTCTTT 306
 DB 250 CTTGATGTCAGCTCCAGTGGATATGCAATTTCCCTTGTATGGAGGACTACTTTTC 309
 QY 307 GCAAACTATGCTTCAAGGGTATGTGACCATCTTAGACCGGTTTCAGCAAAATCTAT 366
 DB 310 GCAAGAAATGCGAGAGAGGATATATTAATGCTCGATCTTTCAGCACAATAT 369
 QY 367 GGAAGACGATGGCGGACTCTCTGTTTATTTCTGCTGACTGATGGGAGAAATGTTCTGGCT 426
 DB 370 GGCACAGATCGTGGCTGATGATGTTTCCAGCACTTCTTGGTGAACATTTCTGGACA 429
 QY 427 GCAGCAATTTTCTGCTTGGGAGCCACATCAGCGTGTATCGATGATGATATGCAC 486
 DB 430 GCAGCAATTTTCTGCGGACATTTGGTGAACACTGTGCGTAATTTTGAATCGACATGAAT 489
 QY 487 ATTTCTGTCATCTCTGCACTCATGTCCACTCTGTACACACTTGTGGAGGCTCTAT 546
 DB 490 GCATGATGACCTCTGCGGCTGATTTGCGGCTATTTCTACACATTCACCGGTGGATATAT 549
 QY 547 TCTGTCGCTACACTGATGCTGCTGCTCTTTGCTGATTTTGTAGGCTCTGATCAGC 606
 DB 550 GCATGCGGTACACTGACGCTGCTTCACTATTTTGCATTTTTCGCTGCTGCTGCTTTC 609
 QY 607 GTCCCTTTTGCATTTGTCATCTCTGCACTCATGTCCAGATCGGCTGCTGCTGCTGCTG 666
 DB 610 GTGCGCGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 661
 QY 667 AAATACCAAAAGCGGTGGCTGGAACTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 726
 DB 662 -----CGGACTGGATGGAGAGATTGGAGGATTTCAAGAAACATCTCTCTGGAATGAT 714
 QY 727 AGTTTCTGTTGTTGATGCTGGTGGTAAATCCCATCGCAGCATCTTTTCAGAGGTTCTFC 786
 DB 715 TGCATGCTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTC 774

QY 787 TCTTCTTCTCTAGCCACCTATCTCAAGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 846
 DB 775 TCTTCT 834
 QY 847 GTGATGGCCATCCAGCCATCTCAATTTGGGGCCATTTGGAGCATCAACACACTGGAACAG 906
 DB 835 CTCATGGCATTTCCACAGCGTTGATCGGTGCAATTTGCCAGGACACAGACTGGAGAATG 894
 QY 907 ACTGATATATTTCCCATCTGGAACAATGGAATGGAATGGAATGGAATGGAATGGAATG 948
 DB 895 ACTGATATATTTCCCATCTGGAACAATGGAATGGAATGGAATGGAATGGAATGGAATG 954
 QY 949 GACATGATTTTACCAATTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1008
 DB 955 AACTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1014
 QY 1009 CTTGGTGCAGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1068
 DB 1015 CTGCGCGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1074
 QY 1069 TCCATGTTTGCACGGAACATCTACAGCTTTTCTTCTAGACAAATGCTTTCGACAAAGAA 1128
 DB 1075 TCAATGTTTGTCTACACATCTGGAAGCTCAAAATTCGCCCTCACGCTCTGAAAAGAA 1134
 QY 1129 ATCGTTTGGGTTATGCGAATCAGAGTGTGTTGTTGGAGCATCTGCAACAGCCATGGCC 1188
 DB 1135 GTGATAATTTGATGAGAATAGCCATCATCTGTTGTTGATCATGCGAACCATCATGGCA 1194
 QY 1189 TTGCTGACGAAACTGTGTATGGCTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1248
 DB 1195 CTTACATTCATCCATCTATGCGCTTGGTATCTTTGTCAGATTTGCTCTACGTCATA 1254
 QY 1249 ATCTTCCCGCAGCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1308
 DB 1255 CTCCTCCCTCAACTATATGTTGTTGTTATATATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1314
 QY 1309 GCAGGTTATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1368
 DB 1315 GCTGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1374
 QY 1369 CAGCCCTTGTATCTTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1428
 DB 1375 CCAGCTCTTCCATTTACCAATGTATACGATGGG-----TACAGTAT 1419
 QY 1429 TTTCCATTTAAACACTTGCATGCTTACATCATCTTAAACCAACATTTGCATCTCTAT 1488
 DB 1420 TTCCCATTTACGAGCACTGCTATGTTATCTTCAATGCTACTATCTATCTATCAATA 1479
 QY 1489 CTAGCCAAATGATCTATTTGAAAGTGGAACTTCCACCTTAAATAGATGATTTGATGCT 1548
 DB 1480 CAATCGGAGAGCTGTTCAAAATCGGAGCTTTGTTCTCCGGAGTGGGAGCTAATGGTGT 1539
 QY 1549 GTTGT 1553
 DB 1540 GTAGT 1544

RESULT 8
 AAD02461
 ID AAD02461 standard; cdna; 386 BP.
 XX AAD02461;
 AC AAD02461;
 XX 24-APR-2001 (first entry)
 DT Mouse p4P6B1 cdna fragment.
 DE Mouse; OMA protein; obese mice adipocyte; p4P6B1;
 KW fuel metabolism disorder; therapy; obesity; diabetes; gene therapy;
 KW anorectic; antidiabetic; ss.
 XX Mus sp.
 OS

QY	1505	TTGAAGTGGAACTTCGCACCTAAAATAGATGATTATTGATGCGTGTTCT	1555	
Db	1347	TTAGTGCCGGCAAGTTCCGCCAGCTACGACTACTTCCGCTGTGTGGT	1395	
	 RESULT 10 ABK73210 ID ABK73210 standard; DNA; 1461 BP. AC AC XX XX ABK73210; XX XX DT DT 13-AUG-2002 (first entry) XX XX DE DE Bacillus licheniformis genomic sequence tag (GST) #501. XX XX XX Differential gene expression; genomic sequenced tag; GST; KW altered culture condition; environmental stress; KW physiological provocation; ds. KW KW OS OS Bacillus licheniformis. XX XX WO200229113-A2. PN PN XX XX PD PD 11-APR-2002.. XX XX XX 05-OCT-2001; 2001WO-US31437. PF PF XX 06-OCT-2000; 2000US-0680598. PR PR XX 27-MAR-2001; 2001US-279526P. PR PR XX (NOVO) NOVOZYMES BIOTECH INC. PA PA PA (NOVO) NOVOZYMES AS. XX XX PI Berka R, Clausen IG; XX XX XX WPI; 2002-416684/44. DR DR XX Monitoring differential expression of several genes in first XX cell relative to expression of same genes in one or more s PT Bacillus cells, by using substrate containing Bacillus gen PT sequenced tag array _ XX PS Claim 4; SEQ ID NO 501; 200pp; English. XX XX The invention describes a method of monitoring differentia XX genes in a first Bacillus cell relative to expression of XX other Bacillus cells, comprising hybridising labelled nuc CC isolated from Bacillus cells to a substrate containing ar CC genomic sequenced tags (GST), examining the array, and de CC relative gene expression by an observed hybridisation rep CC a spot in the array. The method is useful for measuring t CC genes in a first Bacillus cell relative to expression of CC in one or more second Bacillus cells. The method is usefu CC global expression of several genes from a Bacillus cell, CC genes, identifying possible functions of unknown open rea CC monitoring gene copy number variation and stability. Monit CC in expression of genes may be used to provide a represent CC in which Bacillus cells adapt to changes in culture condi CC environmental stress or other physiological provocation CC follow-up characterisation is unnecessary, when one spot CC equals one gene or one open reading frame, since sequenc CC available. This sequence represents a genomic sequence t CC the method of the invention. CC Note: The sequence data for this patent did not form part CC specification, but was obtained in electronic format dir CC at CC ftp.wipo.int/pub/published_pct_sequences.			
XX	Sequence	1461 BP; 315 A; 355 C; 382 G; 409 T; 0 other;		
SQ	Query Match	13.9%; Score 242.6; DB 24; Length		
	Best Local Similarity	52.4%; Pred. No. 3.7e-60;		

Sequence 1461 BP; 315 A; 355 C; 382 G; 400 T; 429 A; 458 G; 487 A; 516 G; 545 G; 574 G; 603 G; 632 G; 661 G; 690 G; 719 G; 748 G; 777 G; 806 G; 835 G; 864 G; 893 G; 922 G; 951 G; 980 G; 1009 G; 1038 G; 1067 G; 1096 G; 1125 G; 1154 G; 1183 G; 1212 G; 1241 G; 1270 G; 1299 G; 1328 G; 1357 G; 1386 G; 1415 G; 1444 G; 1473 G; 1502 G; 1531 G; 1560 G; 1589 G; 1618 G; 1647 G; 1676 G; 1705 G; 1734 G; 1763 G; 1792 G; 1821 G; 1850 G; 1879 G; 1908 G; 1937 G; 1966 G; 1995 G; 2024 G; 2053 G; 2082 G; 2111 G; 2140 G; 2169 G; 2198 G; 2227 G; 2256 G; 2285 G; 2314 G; 2343 G; 2372 G; 2401 G; 2430 G; 2459 G; 2488 G; 2517 G; 2546 G; 2575 G; 2604 G; 2633 G; 2662 G; 2691 G; 2720 G; 2749 G; 2778 G; 2807 G; 2836 G; 2865 G; 2894 G; 2923 G; 2952 G; 2981 G; 3010 G; 3039 G; 3068 G; 3097 G; 3126 G; 3155 G; 3184 G; 3213 G; 3242 G; 3271 G; 3300 G; 3329 G; 3358 G; 3387 G; 3416 G; 3445 G; 3474 G; 3503 G; 3532 G; 3561 G; 3590 G; 3619 G; 3648 G; 3677 G; 3706 G; 3735 G; 3764 G; 3793 G; 3822 G; 3851 G; 3880 G; 3909 G; 3938 G; 3967 G; 3996 G; 4025 G; 4054 G; 4083 G; 4112 G; 4141 G; 4170 G; 4199 G; 4228 G; 4257 G; 4286 G; 4315 G; 4344 G; 4373 G; 4402 G; 4431 G; 4460 G; 4489 G; 4518 G; 4547 G; 4576 G; 4605 G; 4634 G; 4663 G; 4692 G; 4721 G; 4750 G; 4779 G; 4808 G; 4837 G; 4866 G; 4895 G; 4924 G; 4953 G; 4982 G; 5011 G; 5040 G; 5069 G; 5098 G; 5127 G; 5156 G; 5185 G; 5214 G; 5243 G; 5272 G; 5301 G; 5330 G; 5359 G; 5388 G; 5417 G; 5446 G; 5475 G; 5504 G; 5533 G; 5562 G; 5591 G; 5620 G; 5649 G; 5678 G; 5707 G; 5736 G; 5765 G; 5794 G; 5823 G; 5852 G; 5881 G; 5910 G; 5939 G; 5968 G; 5997 G; 6026 G; 6055 G; 6084 G; 6113 G; 6142 G; 6171 G; 6200 G; 6229 G; 6258 G; 6287 G; 6316 G; 6345 G; 6374 G; 6403 G; 6432 G; 6461 G; 6490 G; 6519 G; 6548 G; 6577 G; 6606 G; 6635 G; 6664 G; 6693 G; 6722 G; 6751 G; 6780 G; 6809 G; 6838 G; 6867 G; 6896 G; 6925 G; 6954 G; 6983 G; 7012 G; 7041 G; 7070 G; 7099 G; 7128 G; 7157 G; 7186 G; 7215 G; 7244 G; 7273 G; 7302 G; 7331 G; 7360 G; 7389 G; 7418 G; 7447 G; 7476 G; 7505 G; 7534 G; 7563 G; 7592 G; 7621 G; 7650 G; 7679 G; 7708 G; 7737 G; 7766 G; 7795 G; 7824 G; 7853 G; 7882 G; 7911 G; 7940 G; 7969 G; 7998 G; 8027 G; 8056 G; 8085 G; 8114 G; 8143 G; 8172 G; 8201 G; 8230 G; 8259 G; 8288 G; 8317 G; 8346 G; 8375 G; 8404 G; 8433 G; 8462 G; 8491 G; 8520 G; 8549 G; 8578 G; 8607 G; 8636 G; 8665 G; 8694 G; 8723 G; 8752 G; 8781 G; 8810 G; 8839 G; 8868 G; 8897 G; 8926 G; 8955 G; 8984 G; 9013 G; 9042 G; 9071 G; 9100 G; 9129 G; 9158 G; 9187 G; 9216 G; 9245 G; 9274 G; 9303 G; 9332 G; 9361 G; 9390 G; 9419 G; 9448 G; 9477 G; 9506 G; 9535 G; 9564 G; 9593 G; 9622 G; 9651 G; 9680 G; 9709 G; 9738 G; 9767 G; 9796 G; 9825 G; 9854 G; 9883 G; 9912 G; 9941 G; 9970 G; 10000 G; 10029 G; 10058 G; 10087 G; 10116 G; 10145 G; 10174 G; 10203 G; 10232 G; 10261 G; 10290 G; 10319 G; 10348 G; 10377 G; 10406 G; 10435 G; 10464 G; 10493 G; 10522 G; 10551 G; 10580 G; 10609 G; 10638 G; 10667 G; 10696 G; 10725 G; 10754 G; 10783 G; 10812 G; 10841 G; 10870 G; 10899 G; 10928 G; 10957 G; 10986 G; 11015 G; 11044 G; 11073 G; 11102 G; 11131 G; 11160 G; 11189 G; 11218 G; 11247 G; 11276 G; 11305 G; 11334 G; 11363 G; 11392 G; 11421 G; 11450 G; 11479 G; 11508 G; 11537 G; 11566 G; 11595 G; 11624 G; 11653 G; 11682 G; 11711 G; 11740 G; 11769 G; 11798 G; 11827 G; 11856 G; 11885 G; 11914 G; 11943 G; 11972 G; 12001 G; 12030 G; 12059 G; 12088 G; 12117 G; 12146 G; 12175 G; 12204 G; 12233 G; 12262 G; 12291 G; 12320 G; 12349 G; 12378 G; 12407 G; 12436 G; 12465 G; 12494 G; 12523 G; 12552 G; 12581 G; 12610 G; 12639 G; 12668 G; 12697 G; 12726 G; 12755 G; 12784 G; 12813 G; 12842 G; 12871 G; 12900 G; 12929 G; 12958 G; 12987 G; 13016 G; 13045 G; 13074 G; 13103 G; 13132 G; 13161 G; 13190 G; 13219 G; 13248 G; 13277 G; 13306 G; 13335 G; 13364 G; 13393 G; 13422 G; 13451 G; 13480 G; 13509 G; 13538 G; 13567 G; 13596 G; 13625 G; 13654 G; 13683 G; 13712 G; 13741 G; 13770 G; 13799 G; 13828 G; 13857 G; 13886 G; 13915 G; 13944 G; 13973 G; 14002 G; 14031 G; 14060 G; 14089 G; 14118 G; 14147 G; 14176 G; 14205 G; 14234 G; 14263 G; 14292 G; 14321 G; 14350 G; 14379 G; 14408 G; 14437 G; 14466 G; 14495 G; 14524 G; 14553 G; 14582 G; 14611 G; 14640 G; 14669 G; 14698 G; 14727 G; 14756 G; 14785 G; 14814 G; 14843 G; 14872 G; 14901 G; 14930 G; 14959 G; 14988 G; 15017 G; 15046 G; 15075 G; 15104 G; 15133 G; 15162 G; 15191 G; 15220 G; 15249 G; 15278 G; 15307 G; 15336 G; 15365 G; 15394 G; 15423 G; 15452 G; 15481 G; 15510 G; 15539 G; 15568 G; 15597 G; 15626 G; 15655 G; 15684 G; 15713 G; 15742 G; 15771 G; 15800 G; 15829 G; 15858 G; 15887 G; 15916 G; 15945 G; 15974 G; 16003 G; 16032 G; 16061 G; 16090 G; 16119 G; 16148 G; 16177 G; 16206 G; 16235 G; 16264 G; 16293 G; 16322 G; 16351 G; 16380 G; 164

```

lished_pct_sequences.
5 A; 355 C; 382 G; 409 T; 0 other;
13.9%; Score 242.6; DB 24; Length 1461;
52.4%; pred. No. 3.7e-60;

```


Tue Apr 1 13:57:24 2003

us-10-069-541-5.rng

320 GTTCAAGAGGGTATGTGACCATGTTAGACCCGTTTTCAGCAAAATCTATGAAACACGATGG 379
 2847 GCAAGCAGGGTTACATACCATGTTGGATCCGTTGCAGGATTCCTTTGGTGGCGATGG 2788
 380 GGGGACTCTGTTTATCTGCTGCTGATGGGAGAAATGTTCTGGGTGAGCAATTTCT 439
 2787 GAGGATGTTCTTCTTCCCTGCGCTCTATGCGTGGTCTTTTGGGACCGGATCCTGG 2728
 440 CTGCTTTGGGAGCACCACATCAGCGTGAATCATGATGGGATATGACATTTCTGTCTATCA 499
 2727 CTGCACTTGGCGCACTCTATCGTGTATCATCGACATGGGATACCGGACCTCGGTGATCC 2668
 500 TCTCTGCACTCATTCGCACTCTGTACACACTGTGTGGGAGGCGCTCTATCTGTGGCCTACA 559
 2667 TGTCTCTCTGCACTGCGCATCTTCTACACACTGTGTGGGAGGCGCTCTATCTGTGGCCTACA 2608
 560 CTGATGCTGTTTCACTCTTTTGGCAATTTTGTAGGCTGTGGATCAGCGTCCCTTTTGCAT 619
 2607 CGGACGTGATCCAGTGTCTTCTGCACTCTTCTACACACTGTGTGGGAGGCGCTCTATCTGTGGCCT 2548
 620 TGTACATCTCTGCACTGCGACATCGGGTTCACCTGCTGTGATGCCAAATACCAAAAGC 679
 2547 GGAGCAAGCAGCAGTGGGCGCCTGAGTACCTGGAGGTGGAT----- 2504
 680 CGTGGCTGGGAACTGTTGACTCATCTGAAGTCTACTCTTGGCTTGATAGTTTCTGTGTGT 739
 2503 --TGGATTGGGCACTGGGAGCCTAAAGACATTTGGCTGTACATAGACTACGCTTGTCTG 2446
 740 TGATGCTGGTGGAACTCCCATGGCAAGCATATTTTCAGAGGTTCTCTTCTTCTTCTCAG 799
 2445 TCGTCTTGGTGGCACTCCCTGGCAGGTCTACTTCCAGCGGTGCTGTCCAGCAAAACGG 2386
 800 CCACCTATGCTCAAGTGTCTCTCTCCCTGGCAGCTTTCCGGTGGCTGTGATGGCCATCC 859
 2385 CAGGAAGGCGCCAGCTTCTCTCTATGTTGACCGCGGATGCAATTTGATGGCCATTC 2326
 860 CAGCCATACATCTTGGGCGCATG 883
 2325 CCGCGTGTCTATCGGAGCGATTG 2302

RESULT 13
 AAS82193
 ID AAS82193 standard; CDNA; 1094 BP.
 XX
 AC AAS82193;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #17997.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG18006.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in

9892 ATGGCTTTCCATGTGGAAGGACTGATAGTATCATCTGTTTCTACTCTTAATTTTGGTG 9951
 61 GTTGAATATGGCTGCTGGAGACCAACAAACAGTGGCAGCGCAGAGCGCAGCGAA 120
 9952 GTTGAATATGGCTGCTGGAGACCAACAAACAGTGGCAGCGCAGAGCGCAGCGAA 10011
 121 GCATCATATGTTGGTGGCGGAGATATTTGTTTATTTGTTGGTGGATTTACCATGACAGCT 180
 10012 GCATCATATGTTGGTGGCGGAGATATTTGTTTATTTGTTGGTGGATTTACCATGACAGCT 10071
 181 ACCT 184
 10072 ACGT 10075

RESULT 12
 ABL29568/c
 ID ABL29568 standard; DNA; 4223 BP.
 XX
 AC ABL29568;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 40177.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 40177; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 4223 BP; 1128 A; 1024 C; 964 G; 1107 T; 0 other;
 Query Match 10.3%; Score 179.6; DB 23; Length 4223;
 Best Local Similarity 57.2%; Pred. No. 1.8e-41;
 Matches 357; Conservative 0; Mismatches 249; Indels 18; Gaps 1;
 260 CACCAATGGATATCTCTTACTCTGATTTTAGTGGCGCTGTTCTTTGCAAAACCTATGC 319
 2907 CAAAAGTGATTCTATCCCTCTCCAAACTTAGTGGCATCTCTTTGCAATCCCATGC 2848

PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 XX biodiversity -
 XX
 XX Claim 1: SEQ ID No 17997; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX SQ Sequence 1094 BP; 325 A; 210 C; 267 G; 292 T; 0 other;

Query Match 9.6%; Score 167.6; DB 23; Length 1094;
 Best Local Similarity 96.8%; Pred. No. 2.8e-38;
 Matches 182; Conservative 0; Mismatches 4; Indels 2; Gaps 1;
 QY 1 ATGGCTTTCATGTGAAGACTAGTACATCATCGTGTTCACCTTCTAATTTGCTG 60
 Db 1 ATGGCTTTCATGTGAAGACTAGTACATCATCGTGTTCACCTTCTAATTTGCTG 60
 QY 61 GTTGGAAATATGGCTGCCTGGAGACCAAAACAGTGGCAGCGAGAGCGCAGCGAA 120
 Db 61 GTTGGAAATATGGCTGCCTGGAGACCAAAACAGTGGCAGCGAGAGCGCAGCGAA 118
 QY 121 GCCATCATAGTTGGTGGCGGAGATATGGTTATTGGTGGTGGATTACCATGACAGCT 180
 Db 119 GCCATCATAGTTGGTGGCGGAGATATGGTTATTGGTGGTGGATTACCATGACAGAT 178
 QY 181 ACCTGGGT 188
 Db 179 ACATATGT 186

RESULT 14
 AAH49203
 ID AAH49203 standard; DNA; 240 BP.
 XX
 AC AAH49203;
 XX
 XX 26-NOV-2001 (first entry)
 DE Human CHOT exon 3 region DNA.
 XX
 KW CHOT; human; choline transporter; chromosome 2q11-13; neurotropic;
 KW neuroprotective; gene therapy; antisense therapy; degenerative disease;
 KW cognitive disorder; Alzheimer's disease; ds.
 XX
 OS Homo sapiens.
 XX
 XX DE10009055-A1.
 XX
 XX 30-AUG-2001.
 PD
 XX 28-FEB-2000; 2000DE-1009055.
 PF
 XX

PR 28-FEB-2000; 2000DE-1009055.
 XX (BRUE/) BRUESS M.
 PA (BOEN/) BOENISCH H.
 XX
 PI Bruess M, Boenisch H;
 XX WPI; 2001-590709/67.
 DR
 XX A new gene encoding human choline transporter, designated hCHOT is
 PT located on chromosome 2q11-13 and is useful to treat degenerative
 PT disorders such as Alzheimer's disease -
 XX
 PS Disclosure; Page 7; 12pp; German.
 XX
 CC This invention describes a novel gene encoding human choline transporter,
 CC designated hCHOT which is located on chromosome 2q11-13. The products of
 CC the invention have neurotropic and neuroprotective activity and can be used
 CC for gene or antisense therapy. (I) is used to treat degenerative disease,
 CC particularly cognitive disorders such as Alzheimer's disease. Sense and
 CC antisense oligonucleotides derived from the gene may be used in
 CC diagnostics and other techniques. This sequence represents the exon 3
 CC fragment encoding the human CHOT protein described in the invention.
 XX
 SQ Sequence 240 BP; 60 A; 45 C; 53 G; 82 T; 0 other;
 Query Match 9.4%; Score 163; DB 22; Length 240;
 Best Local Similarity 100.0%; Pred. No. 2.7e-37;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 287 TTTTAGTGGCCCTGTTCTTTGCAAAACCTATGCGTTCAAAAGGGTATGTGACCATGTAG 346
 Db 60 TTTTAGTGGCCCTGTTCTTTGCAAAACCTATGCGTTCAAAAGGGTATGTGACCATGTAG 119
 QY 347 ACCGTTTCAGCAAAATCTATGAAACGATGGCGGACCTCTCTTTATTCCTGCACTGA 406
 Db 120 ACCGTTTCAGCAAAATCTATGAAACGATGGCGGACCTCTCTTTATTCCTGCACTGA 179
 QY 407 TGGGAGAAATCTCTGGGCTGCGAGCAATTTCTCTCTTTGGG 449
 Db 180 TGGGAGAAATCTCTGGGCTGCGAGCAATTTCTCTCTTTGGG 222
 RESULT 15
 ABA53620
 ID ABA53620 standard; DNA; 455 BP.
 XX
 AC ABA53620;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human foetal liver single exon nucleic acid probe #1925.
 DE
 XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 KW
 XX Homo sapiens.
 OS
 XX WO200157277-A2.
 PN
 XX 09-AUG-2001.
 PD
 XX 30-JAN-2001; 2001WO-US00669.
 PF
 XX
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA

us-10-069-541-5.rng

Tue Apr 1 13:57:24 2003

```

PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
PT Claim 1; SEQ ID NO 1925; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 455 BP; 121 A; 108 C; 85 G; 141 T; 0 other;
XX
Query Match 8.9%; Score 155; DB 22; Length 455;
Best Local Similarity 100.0%; Pred. No. 8.3e-35;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 741 GATGCTGGGTGGAATCCCATGGCAAGCATACTTCAGAGGGTTCTCTCTTCCTCAGC 800
DB 266 GATGCTGGGTGGAATCCCATGGCAAGCATACTTCAGAGGGTTCTCTCTTCCTCAGC 325
QY 801 CACCTATGCTCAAGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 860
DB 326 CACCTATGCTCAAGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 385
QY 861 AGCCATCACTATGGGGCCATTGGAGCATCAACAG 895
DB 386 AGCCATCACTATGGGGCCATTGGAGCATCAACAG 420

```

Search completed: March 31, 2003, 16:33:42
Job time : 436 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2003, 15:57:11 ; Search time 4568 seconds
(without alignments)
11104.883 Million cell updates/sec

Title: US-10-069-541-5
Perfect score: 1743
Sequence: 1 atggctttccatgtggaagg.....ctgaagataattcacagtga 1743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_ph.*
7: gb_pl.*
8: gb_pr.*
9: gb_ro.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_man.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	1743	100.0	1743	6	BD012719	BD012719 High-affi
2	1743	100.0	1743	9	AF276871	AF276871 Homo sapi
3	1743	100.0	1743	23	BD005267	BD005267 High-affi
4	1743	100.0	1813	9	HS401466	AJ401466 Homo sapi
5	1743	100.0	5158	9	AB043997	AB043997 Homo sapi
6	1394.2	80.0	1743	6	BD012718	BD012718 High-affi
7	1394.2	80.0	1743	23	BD005266	BD005266 High-affi
8	1394.2	80.0	4904	10	AB030947	AB030947 Rattus no
9	1375	78.9	1743	10	AF276872	AF276872 Mus muscu
10	1373.4	78.8	1743	6	BD012720	BD012720 High-affi
11	1373.4	78.8	1743	23	BD005268	BD005268 High-affi
12	1373.4	78.8	4938	6	AX080443	AX080443 Sequence
13	1367	78.4	1743	10	MMU401467	AJ401467 Mus muscu
14	867	49.7	2528	5	TMA420808	AJ420808 Torpedo m
15	630.8	36.2	2239	9	HS4308384	AJ308384 Homo sapi
16	630.8	36.2	190043	9	AC009963	AC009963 Homo sapi
17	502.8	28.8	17479	2	AC106657	AC106657 Rattus no
18	501.2	28.8	155131	2	AC102873	AC102873 Mus muscu
19	431.4	24.8	3326	3	AY011119	AY011119 Limulus p
20	405.8	23.3	3255	3	AY047521	AY047521 Drosophil
21	363.8	20.9	1731	6	BD012717	BD012717 High-affi
22	363.8	20.9	1731	23	BD005265	BD005265 High-affi
23	363.8	20.9	1985	3	AB030946	AB030946 Caenorhab
24	279.6	16.0	386	6	AX080449	AX080449 Sequence
25	242.6	13.9	1461	6	AX432086	AX432086 Sequence
26	226	13.0	1657	9	HS4308383	AJ308383 Homo sapi
27	179.6	10.3	1178	9	HS4308378	AJ308378 Homo sapi
28	179.6	10.3	186989	3	AC007812	AC007812 Drosophil
29	179.6	10.3	189117	3	AC009395	AC009395 Drosophil
30	179.6	10.3	255624	2	AE003723	AE003723 Drosophil
31	167.6	9.6	140156	2	AC017381	AC017381 Drosophil
32	163	9.4	2326	9	HS4308379	AJ308379 Homo sapi
33	155	8.9	1467	9	HS4308382	AJ308382 Homo sapi
34	151.8	8.7	40993	3	CBG45E19	AC084631 Caenorhab
35	150.2	8.6	736	9	HS4308381	AJ308381 Homo sapi
36	150	8.6	1308	9	HS4308380	AJ308380 Homo sapi
37	141.6	8.1	39908	3	CEC48D1	Z81049 Caenorhabdi
38	141.6	8.1	330724	2	CEY67H2	AL022475 Caenorhab
39	132	7.6	152021	2	AC010923	AC010923 Drosophil
40	109	6.3	17479	2	AC106657	AC106657 Rattus no
41	51.8	3.0	10732	6	E32986	E32986 Gene encodi
42	48.2	2.8	2781	1	BSU92466	U92466 Bacillus su
43	48.2	2.8	3933	1	AF011545	AF011545 Bacillus
44	48.2	2.8	213190	1	BSUR0004	Z99107 Bacillus su
45	48	2.8	2426	4	SSC487855	AJ487855 Sus scrof

ALIGNMENTS

RESULT 1
BD012719
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

BD012719
High-affinity choline transporter.
BD012719, 1 GI:22092908
WO 0116315-A/3.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1743)
Haga,T. and Okuda,T.
High-affinity choline transporter
Patent: WO 0116315-A 3 08-MAR-2001;

BD012719 1743 bp DNA linear PAT 02-AUG-2002

Pred. No. is the number of results predicted by chance to have a

JAPAN SCIENCE AND TECHNOLOGY CORP, TATSUYA HAGA, TAKASHI OKUDA
OS Homo sapiens (human)
PN WO 0116315-A/3
PD 08-MAR-2001
PR 18-AUG-1999 JP 99P 240642.27-DEC-1999 JP 99P 368991 PI
TATSUYA HAGA, TAKASHI OKUDA
PC C12N15/12, C07K14/47, C12Q1/68, C07K19/00, C07K16/18, C12N5/10, PC
A61K38/17,
A61K45/00, A61P25/28, G01N33/53, A01K67/027
CC
FH Key Location/Qualifiers
FT CDS (1)..(1743).
FEATURES
source Location/Qualifiers
I. 1743
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 412 a 393 c 406 g 532 t
ORIGIN
Query Match 100.0%; Score 1743; DB 6; Length 1743;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCTTTCCATGTGGAGGACTGATAGCTATCATCGTGTCTACCTTCTAATTTGCTG 60
Db 1 ATGGCTTTCCATGTGGAGGACTGATAGCTATCATCGTGTCTACCTTCTAATTTGCTG 60
QY 61 GTTGGAAATATGGCTGCTGGAGAACCAAAACAGTGGCAGCGAGAGCGCAGCGAA 120
Db 61 GTTGGAAATATGGCTGCTGGAGAACCAAAACAGTGGCAGCGAGAGCGCAGCGAA 120
QY 121 GCCATCATAGTTGGTGGCCGAGATATTGGTTATTGGTTGGTGGATTTACCATGACAGCT 180
Db 121 GCCATCATAGTTGGTGGCCGAGATATTGGTTATTGGTTGGTGGATTTACCATGACAGCT 180
QY 181 ACCTGGTGGAGGAGGTATATCAATGGACACCTGAAGCAGTTTATGTACAGGTTAT 240
Db 181 ACCTGGTGGAGGAGGTATATCAATGGACACCTGAAGCAGTTTATGTACAGGTTAT 240
QY 241 GGCTAGCTTGGGCTCAGGACCAATTTGGATATCTCTAGTCTGATTTAGTGGGCGTG 300
Db 241 GGCTAGCTTGGGCTCAGGACCAATTTGGATATCTCTAGTCTGATTTAGTGGGCGTG 300
QY 301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
Db 301 TTCTTTGCAAAACCTATGCGTTCAAAGGGGTATGTGACCATGTTAGACCCGTTTCAGCAA 360
QY 361 ATCTATGGAAACGCGATGGGCGGACTCCTGTTATTCTCGACCTGATGGGAGAAATGTC 420
Db 361 ATCTATGGAAACGCGATGGGCGGACTCCTGTTATTCTCGACCTGATGGGAGAAATGTC 420
QY 421 TGGGCTGCAGCAATTTCTCTGCTTTGGGAGCCACCATGACGCTGATCGATGGAT 480
Db 421 TGGGCTGCAGCAATTTCTCTGCTTTGGGAGCCACCATGACGCTGATCGATGGAT 480
QY 481 ATGCACATTTCTGTCATCATCTGCACTCATTTGCCACTCTGTACACACTGGTGGGAGG 540
Db 481 ATGCACATTTCTGTCATCATCTGCACTCATTTGCCACTCTGTACACACTGGTGGGAGG 540
QY 541 CTCATCTGTGGCCCTACACTGATGCTGCTGAGTCTTTGCAATTTTGTAGGCGTGTG 600
Db 541 CTCATCTGTGGCCCTACACTGATGCTGCTGAGTCTTTGCAATTTTGTAGGCGTGTG 600
QY 601 ATCAGCGTCCCTTTGCAATTTGTCATCTCTGAGTCCGACGATCGGCTTACCTGCTGTG 660
Db 601 ATCAGCGTCCCTTTGCAATTTGTCATCTCTGAGTCCGACGATCGGCTTACCTGCTGTG 660
QY 661 CATGCCAAATACCAAAGCGGTGGGAACTGTTGACTCATCTGAAGTCTACTCTTTG 720
Db 661 CATGCCAAATACCAAAGCGGTGGGAACTGTTGACTCATCTGAAGTCTACTCTTTG 720
QY 721 CTTGATAGTTTTCTGTTGTTGCTGGGTGGAAATCCCATGGCAGCATACTTTCAGAGG 780

Db 721 CTTGATAGTTTTCTGTTGTTGATGCTGGTGGAAATCCCATGGCAAGCATACTTTTCAGAGG 780
QY 781 GTTCTCTCTTCTTCTCCTCAGCCACTATGCTCAAGTGTCTTCTTCTTGGCAGCTTTTCGGG 840
Db 781 GTTCTCTCTTCTTCTCCTCAGCCACTATGCTCAAGTGTCTTCTTCTTGGCAGCTTTTCGGG 840
QY 841 TGCTGTGATGGCCATCCCGAGCATACTCATTTGGGGCCATTGGAGCATCAACAGACTGG 900
Db 841 TGCTGTGATGGCCATCCCGAGCATACTCATTTGGGGCCATTGGAGCATCAACAGACTGG 900
QY 901 AACGACATGTCATATGGGCTTCCAGATGCCAAGACTACAGAAGAGGAGAGCATGATTTTA 960
Db 901 AACGACATGTCATATGGGCTTCCAGATGCCAAGACTACAGAAGAGGAGAGCATGATTTTA 960
QY 961 CCAATTTGTTCTGCAGTATCTCTGCCCTGTGTATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
Db 961 CCAATTTGTTCTGCAGTATCTCTGCCCTGTGTATTTCTTTCTTTGGTCTTGGTGCAGTT 1020
QY 1021 TCTGCTGCTGTTATGTCTATCAGCAGATTTCTTCCATCTTGTGACCAAGTTCCTGTTTGA 1080
Db 1021 TCTGCTGCTGTTATGTCTATCAGCAGATTTCTTCCATCTTGTGACCAAGTTCCTGTTTGA 1080
QY 1081 CGGAACATCTACAGCTTTCTTCCAGCAAAATGCTTTCGGACAAAGAAATCGTTTGGGTT 1140
Db 1081 CGGAACATCTACAGCTTTCTTCCAGCAAAATGCTTTCGGACAAAGAAATCGTTTGGGTT 1140
QY 1141 ATGCGAATCACAGTGTGTTGTTGGAGCATCTGCAACAGCCATGGCTTCTGACGAA 1200
Db 1141 ATGCGAATCACAGTGTGTTGTTGGAGCATCTGCAACAGCCATGGCTTCTGACGAA 1200
QY 1201 ACTGTGATGGGCTCTGGTACCTCAGTTCTGAGCTTGTGTTTACATCTGTTATCTTCCCGAG 1260
Db 1201 ACTGTGATGGGCTCTGGTACCTCAGTTCTGAGCTTGTGTTTACATCTGTTATCTTCCCGAG 1260
QY 1261 CTGCTTTGTTGTTACTTCTTTTAAAGGAAACCAACACCTATGGGCGCTGGCAGGTTATGTT 1320
Db 1261 CTGCTTTGTTGTTACTTCTTTTAAAGGAAACCAACACCTATGGGCGCTGGCAGGTTATGTT 1320
QY 1321 TCTGGCTCTTCTCTGAGTAATGCTGAGGGAGCCATATCTGATCTTTCAGCCCTTGATC 1380
Db 1321 TCTGGCTCTTCTCTGAGTAATGCTGAGGGAGCCATATCTGATCTTTCAGCCCTTGATC 1380
QY 1381 TTCTACCTCTGGCTATTACCTGATGATAATGTTATATATATATATATATATATATATATAT 1440
Db 1381 TTCTACCTCTGGCTATTACCTGATGATAATGTTATATATATATATATATATATATATAT 1440
QY 1441 ACACCTTGGCATGGTTACATCATCTTAAACCAACATTTGCAATCTCTATCTAGCCAAAT 1500
Db 1441 ACACCTTGGCATGGTTACATCATCTTAAACCAACATTTGCAATCTCTATCTAGCCAAAT 1500
QY 1501 CTATTTGAAAGTGGAACTTGGCCACCTAAATAGATGTTATGCTGTTTGTGCAAGA 1560
Db 1501 CTATTTGAAAGTGGAACTTGGCCACCTAAATAGATGTTATGCTGTTTGTGCAAGA 1560
QY 1561 CACAGTGAAGAAACATGATGATAAGCAATTTCTGCAAAATGAAATATATATATATATAT 1620
Db 1561 CACAGTGAAGAAACATGATGATAAGCAATTTCTGCAAAATGAAATATATATATATATAT 1620
QY 1621 GAACCTGCACTTGTGAAGCCAGCAGAGCATGACCCCTCAGCTCAACTTTTCACCAATAA 1680
Db 1621 GAACCTGCACTTGTGAAGCCAGCAGAGCATGACCCCTCAGCTCAACTTTTCACCAATAA 1680
QY 1681 GAGGCTCTCTGATGTTGATTTCCAGTCCAGAGGCTCTGGGACTGAAGATAATTTACAG 1740
Db 1681 GAGGCTCTCTGATGTTGATTTCCAGTCCAGAGGCTCTGGGACTGAAGATAATTTACAG 1740
QY 1741 TGA 1743
Db 1741 TGA 1743

RESULT 2

Tue Apr 1 13:57:23 2003

AF276871 1743 bp mRNA linear PRI 27-NOV-2000
 LOCUS Homo sapiens high affinity choline transporter (SLC5A7) mRNA,
 DEFINITION complete cds.
 ACCESSION AF276871
 VERSION AF276871.1 GI:10998441
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1743)
 Apparsundaram, S., Ferguson, S.M., George, A.L. Jr. and Blakely, R.D.
 Molecular cloning of a human, hemicholinium-3-sensitive choline
 transporter.
 Biochem. Biophys. Res. Commun. 276 (3), 862-867 (2000)
 JOURNAL Biochem. Biophys. Res. Commun. 276 (3), 862-867 (2000)
 MEDLINE 20483599
 PUBMED 11027560
 REFERENCE 2 (bases 1 to 1743)
 Apparsundaram, S., Ferguson, S.M. and Blakely, R.D.
 Direct Submission
 Submitted (09-JUN-2000) Department of Pharmacology and Center for
 Molecular Neuroscience, Vanderbilt University, 23rd Avenue South at
 Pierce, Nashville, TN 37232-6420, USA
 JOURNAL Location/Qualifiers
 FEATURES
 source
 1. .1743
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2q12"
 1. .1743
 /gene="SLC5A7"
 1. .1743
 /note="hCHT; solute carrier family 5 member 7"
 /codon_start=1
 /product="high affinity choline transporter"
 /protein_id="AAG25940.1"
 /db_xref="GI:10998442"
 /translation="MAFHVEGLIAIVFYLLILVGLVWAAWRTKNSGABERSALIV
 GGRDILVGGVGTMTATWVGGYINGTAEAVYVPGYGLNAQAPIGYSLILGLGF
 AKPRSGYVMTLDPFOQIYKRMGGLIFIPALMGEMFWAAIFSAIGATISVIDVD
 MHSIVTISALITLYTLVGLYSVATDVOLFCIFVGLWISVPFALSHPAVADIGET
 AVHAKYQWPLGTVDSSVYSLWDSFLLMLGGIPQWAFORVLSSTASATVAQLSEL
 AAFGLVMAIPAILIGALGASTDNQATAYGLPDKPTTEADMLPILVLYLCPVVISF
 FGLGAVSAVMSSADSSILSSASSFARNIYQLSFQNASDKEIYVWVRITVFEVGSAA
 TAMALLTKTVGLWYLSLDLYIYVIFPOLLCVLFVKGNTYICAVAGYVSLFLESLGP
 EPLYLQPLIFPGYIPDNGIYVQKPFKTLAMVTSFLTNICISYLAKYLFESGLTP
 PKLDVFAVARHSEENNDKTIILVKNENIKLDELALVKPROSMTLSSFTTNKEAFLDV
 DSSPEGSCTEDNLO"
 BASE COUNT 412 a 393 c 406 g 532 t
 ORIGIN
 Query Match 100.0%; Score 1743; DB 9; Length 1743;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGGCTTCCATGTGGAAGACTGATAGCTATCATCTGTCTPACCTTCTAATTTGGCTG 60
 DB 1 ATGGCTTCCATGTGGAAGACTGATAGCTATCATCTGTCTPACCTTCTAATTTGGCTG 60
 QY 61 GTTGAATATGGCTGCCGTGGAACCAAAACAGTGGCAGCGCAGAAAGCGCAGAA 120
 DB 61 GTTGAATATGGCTGCCGTGGAACCAAAACAGTGGCAGCGCAGAAAGCGCAGAA 120
 QY 121 GCCATCATAGTTGGTGGCCGAGATATGGTTTATGGTTGGTGGATTTACCATGACAGCT 180
 DB 121 GCCATCATAGTTGGTGGCCGAGATATGGTTTATGGTTGGTGGATTTACCATGACAGCT 180
 QY 181 ACCTGGTCCGAGGGGTATCAATGGCACAGCTCAAGCAGTTTATGTACCAAGGTTAT 240
 DB 181 ACCTGGTCCGAGGGGTATCAATGGCACAGCTCAAGCAGTTTATGTACCAAGGTTAT 240

241 GGCCTAGCTTGGCTCAGCACCAATTGGATATCTCTTAGTCTGATTTAGTGGCCTG 300
 DB 241 GGCCTAGCTTGGCTCAGCACCAATTGGATATCTCTTAGTCTGATTTAGTGGCCTG 300
 QY 301 TTCTTTGCAAAACCTATGGCTTCAAAAGGGATGTGACCAATGTAGCCGTTTACAGCA 360
 DB 301 TTCTTTGCAAAACCTATGGCTTCAAAAGGGATGTGACCAATGTAGCCGTTTACAGCA 360
 QY 361 ATCTATGGAACGATGGCGGACTCTCTTTATCTCTGCTGCTGATGGGAGAAATGTTT 420
 DB 361 ATCTATGGAACGATGGCGGACTCTCTTTATCTCTGCTGCTGATGGGAGAAATGTTT 420
 QY 421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTTG 480
 DB 421 TGGGCTGCAGCAATTTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTTG 480
 QY 481 ATGCACATTTCTGTCATCATCTCTGCACTCATTCGCCACTCTGTACACACTGGTGGGAGG 540
 DB 481 ATGCACATTTCTGTCATCATCTCTGCACTCATTCGCCACTCTGTACACACTGGTGGGAGG 540
 QY 541 CTCTATTCTGCGCTACACTGATGCTGCTTTCAGCTCTTTTGCATTTTGTAGGCGTGG 600
 DB 541 CTCTATTCTGCGCTACACTGATGCTTTCAGCTCTTTTGCATTTTGTAGGCGTGG 600
 QY 601 ATCAGCGTCCCTTTGCAATGTCATCTCGACGTCGCGAGCATCGGCTTCACTGCTGTG 660
 DB 601 ATCAGCGTCCCTTTGCAATGTCATCTCGACGTCGCGAGCATCGGCTTCACTGCTGTG 660
 QY 661 CATGCCAAATACCAAAAGCGTGGTGGCACTGTGTGACTCATCTGAAAGTCTACTTTGG 720
 DB 661 CATGCCAAATACCAAAAGCGTGGTGGCACTGTGTGACTCATCTGAAAGTCTACTTTGG 720
 QY 721 CTGTAGTCTTCTGCTTTCATGCTGGTGGAAATCCCATGCAAGCATACCTTTCAGAGG 780
 DB 721 CTGTAGTCTTCTGCTTTCATGCTGGTGGAAATCCCATGCAAGCATACCTTTCAGAGG 780
 QY 781 GTTCTCTCTTCTCAGCACCTTCTCAAGTGTCTCTCTCTCTGCGACGTTTCGGG 840
 DB 781 GTTCTCTCTTCTCAGCACCTTCTCAAGTGTCTCTCTCTCTGCGACGTTTCGGG 840
 QY 841 TGCCTGGTGTATGGCATCCAGCCATCTATTTGGGGCCATTTGGAGCATCAACAGACTGG 900
 DB 841 TGCCTGGTGTATGGCATCCAGCCATCTATTTGGGGCCATTTGGAGCATCAACAGACTGG 900
 QY 901 AACACAGCTGCATATGGCTTCCAGATCCCAAGCTACAGAGGAGCAGACATGATTTTA 960
 DB 901 AACACAGCTGCATATGGCTTCCAGATCCCAAGCTACAGAGGAGCAGACATGATTTTA 960
 QY 961 CCAATTTGTTCTGCAGTATCTCTGCCCTGTGTATTTCTTTTCTGCTGTGGTGCAGTT 1020
 DB 961 CCAATTTGTTCTGCAGTATCTCTGCCCTGTGTATTTCTTTCTGCTGTGGTGCAGTT 1020
 QY 1021 TCTGCTGCTGTTATGTCATCAGCAGATTTCTCCATCTTGTGTCAGCAAGTTCATGTTTGA 1080
 DB 1021 TCTGCTGCTGTTATGTCATCAGCAGATTTCTCCATCTTGTGTCAGCAAGTTCATGTTTGA 1080
 QY 1081 CGGAACATCTACAGCTTTCTTCCAGACAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
 DB 1081 CGGAACATCTACAGCTTTCTTCCAGACAAATGCTTCGGACAAAGAAATCGTTTGGGTT 1140
 QY 1141 ATCGAATACAGTGTGTTGTTTGGAGCATCTGCAACAGCATGGCTTGTCTGACAGAA 1200
 DB 1141 ATCGAATACAGTGTGTTGTTTGGAGCATCTGCAACAGCATGGCTTGTCTGACAGAA 1200
 QY 1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTCACCTGTTATCTCTCCCCAG 1260
 DB 1201 ACTGTGTATGGGCTCTGGTACCTCAGTTCTGACCTTGTTCACCTGTTATCTCTCCCCAG 1260
 QY 1261 CTGCTTGTGTACTTCTTTTAAAGGAAACCAACACCTATGGGCGCGTGGCAGGTTATGTT 1320
 DB 1261 CTGCTTGTGTACTTCTTTTAAAGGAAACCAACACCTATGGGCGCGTGGCAGGTTATGTT 1320
 QY 1321 TCTGGCCTTCTTCTCTGAGAATAACTGGAGGGAGCCATATCTGTATCTTTCAGCCCTTGTATC 1380

Tue Apr 1 13:57:23 2003

961 CCAATTGCTGAGTATCTGCGCTGTGTATATTTCTTCTTTTGTCTTGTGCGAGTT 1020
 Db CCAATTGCTGAGTATCTGCGCTGTGTATATTTCTTCTTTTGTCTTGTGCGAGTT 1020
 Qy TCTGCTGCTGTATGTCATCAGCAGATCTTCCATCTTGTGTCAGCAAGTTCATGTTGCA 1080
 Db TCTGCTGCTGTATGTCATCAGCAGATCTTCCATCTTGTGTCAGCAAGTTCATGTTGCA 1080
 Qy CGGAACATCTACCACTTTCCTTTCAGCAAAATGCTTCGGAAGAAATCGTTGGGTT 1140
 Db CGGAACATCTACCACTTTCCTTTCAGCAAAATGCTTCGGAAGAAATCGTTGGGTT 1140
 Qy ATGCGAATCAGAGTCTTGTGTTGGAGCATCTGCAACAGCCATGGCTTCGACGAA 1200
 Db ATGCGAATCAGAGTCTTGTGTTGGAGCATCTGCAACAGCCATGGCTTCGACGAA 1200
 Qy ACTGTTGATGGCTCTGTTGTTGAGCACTCTGCAACAGCCATGGCTTCGACGAA 1260
 Db ACTGTTGATGGCTCTGTTGTTGAGCACTCTGCAACAGCCATGGCTTCGACGAA 1260
 Qy CTGCTTTGCTACTCTTCTTAAAGGAACCAACACCTATGGGGCGTGGCAGTTATGTT 1320
 Db CTGCTTTGCTACTCTTCTTAAAGGAACCAACACCTATGGGGCGTGGCAGTTATGTT 1320
 Qy CTGCTTTGCTACTCTTCTTAAAGGAACCAACACCTATGGGGCGTGGCAGTTATGTT 1320
 Db CTGCTTTGCTACTCTTCTTAAAGGAACCAACACCTATGGGGCGTGGCAGTTATGTT 1320
 Qy TCTGGCCTCTTCTGAGCAATTAACCTGAGGGGAGCCATATCTGTATCTTCAAGCCCTTGATC 1380
 Db TCTGGCCTCTTCTGAGCAATTAACCTGAGGGGAGCCATATCTGTATCTTCAAGCCCTTGATC 1380
 Qy TTCTACCTGGCTATTACCTGATGATTAATGTTATATATATATATATATATATATATATAT 1440
 Db TTCTACCTGGCTATTACCTGATGATTAATGTTATATATATATATATATATATATATATAT 1440
 Qy TTCTACCTGGCTATTACCTGATGATTAATGTTATATATATATATATATATATATATATAT 1500
 Db TTCTACCTGGCTATTACCTGATGATTAATGTTATATATATATATATATATATATATATAT 1500
 Qy ACCTTGGCATGGTGTACATATCTTAAACCAATTTGATCTCTCTATCTAGCAAGTAT 1560
 Db ACCTTGGCATGGTGTACATATCTTAAACCAATTTGATCTCTCTATCTAGCAAGTAT 1560
 Qy CTATTGAAAGTGAACCTTGGCACTTAAATAGATGATTTGATGCTGTTGTCAGAA 1620
 Db CTATTGAAAGTGAACCTTGGCACTTAAATAGATGATTTGATGCTGTTGTCAGAA 1620
 Qy CACAGTGAAGAAACATGATGAAGCAATCTTGTCAAAAATGAAATATTAATATAGAT 1680
 Db CACAGTGAAGAAACATGATGAAGCAATCTTGTCAAAAATGAAATATTAATATAGAT 1680
 Qy GAACCTGCACTTGTGAAGCCAGCAGACATGACCTCAGCTCAACTTCAACCAATAA 1740
 Db GAACCTGCACTTGTGAAGCCAGCAGACATGACCTCAGCTCAACTTCAACCAATAA 1740
 Qy GAGGCTCTCTGATGTTGATTTCCAGTCCAGAGGGTCTGGGACTGAAGATATTTACAG 1740
 Db GAGGCTCTCTGATGTTGATTTCCAGTCCAGAGGGTCTGGGACTGAAGATATTTACAG 1740

RESULT 4
 Locus HSA401466 1813 bp mRNA linear PRI 16-AUG-2000
 DEFINITION Homo sapiens mRNA for high affinity choline transporter (CHT1 gene)
 ACCESSION AJ401466
 VERSION AJ401466.1 GI:9843753
 KEYWORDS Ch1 gene; high affinity choline transporter.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1813)
 Wientland, A., Bonisch, H. and Brüss, M.
 TITLE Molecular cloning of the human and murine high affinity choline

transporters and characterization of the human gene-structure
 Unpublished
 2 (bases 1 to 1813)
 Brüss, M.
 Direct Submission
 Submitted (14-AUG-2000) Brüss, M., University of Bonn, Pharmacology
 and Toxicology, Reuter str. 2b, D-53113 Bonn, GERMANY
 Location/Qualifiers
 1..1813
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2q11-13"
 /tissue_type="hypothalamus"
 19..1761
 /gene="CHT1"
 19..1761
 /gene="CHT1"
 /function="sodium- and chloride-dependent reuptake of choline"
 /codon_start=1
 /evidence="experimental"
 /product="high affinity choline transporter"
 /protein_id="CAC03717.1"
 /db_xref="GI:9843753"
 /translation="MAFVSLGILAIIVFVLLILVGIWAARTRKNSGASERSEAIIV
 GGRDILGLVGFMTATWGGYINGTAAYVPGYGLAWAQAIPGYSLILGLLFF
 AKPRSKGYVTMLDPFOQIYKGRNGILFIPALMGEMFAAIFSAIGATISIVLDV
 MHSVITISALITATLYLVGLYVAYDVQLFCIFVGLMISVPSFALSHPAVADIGFT
 AVHAKYKPLGTVDSSEVSWLDSFLLMLGGIPQWQAYFORVLSSSATVAQVLSFL
 AAFCLVMAIPAILIGASTDMNQTAYGLPDKPTTEADMLPTVLYLCVPIVIF
 FGLGVAASVMSADSSILSAISSIFARNIYQLSFQNASDKREIYVWVRIVFVFGASA
 TAMALLTKVYGLNVLSSDLVIVIFPOLLCVLFVNTNTYGAIVAGVSLFLBITGS
 EYLYLQPLIIPGYPDNGIYNOKPFKTLAMVTSELTNICYLAKYLFEGSTLP
 PKLDFDVAVVAHRSSENDKILYKVENIKLDELALVPRQSMTLLSFTTNKEAFLDV
 DSSPEGSGEDNLQ"

BASE COUNT 440 a 406 c 417 g 550 t
 ORIGIN
 Query Match 100.0%; Score 1743; DB 9; Length 1813;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCTTTTCCATGTTGGAAGCACTGATAGCTATCATCTGTTCTACCTTCTTAATTTGCTG 60
 Db 19 ATGCTTTTCCATGTTGGAAGCACTGATAGCTATCATCTGTTCTACCTTCTTAATTTGCTG 78
 Qy 61 GTTGAATATCGCTGCTGGAGAACCAAAACAGTGCAGCGCAGAGCGCAGCGAA 120
 Db 79 GTTGAATATCGCTGCTGGAGAACCAAAACAGTGCAGCGCAGAGCGCAGCGAA 138
 Qy 121 GCCATCATAGTTGGTGGCGGAGATATTTGTTTATTTGTTGGTGGATTTTACCATGACGCT 180
 Db 139 GCCATCATAGTTGGTGGCGGAGATATTTGTTTATTTGTTGGTGGATTTTACCATGACGCT 198
 Qy 181 ACCTGGTGGAGAGGGTATATCAATGGCAGCTGAAGCAGTGTATGTACCAAGTTAT 240
 Db 199 ACCTGGTGGAGAGGGTATATCAATGGCAGCTGAAGCAGTGTATGTACCAAGTTAT 258
 Qy 241 GGCTAGCTTGGGCTCAGCACCAATGATATCTCTTAGTCTGATTTAGGTGGCCTG 300
 Db 259 GGCTAGCTTGGGCTCAGCACCAATGATATCTCTTAGTCTGATTTAGGTGGCCTG 318
 Qy 301 TTCTTTGCAAAACCTATGCTTCAAAAGGGTATGTGACCATGTTAGACCGTTTCAGCA 360
 Db 319 TTCTTTGCAAAACCTATGCTTCAAAAGGGTATGTGACCATGTTAGACCGTTTCAGCA 378
 Qy 361 ATCTATGAAACGATGCGGCGACTCCTCTTTATCTCTGCTACTGATGGAGAAATGTC 420
 Db 379 ATCTATGAAACGATGCGGCGACTCCTCTTTATCTCTGCTACTGATGGAGAAATGTC 438
 Qy 421 TGGGCTGCGCAATTTCTCTGCTTTGGGACCAACATCAGCGTGATCATGCTGAT 480

Db 439 TGGGCTGACGCAATTTCTCTGCTTTGGGAGCCACCATCAGCGTGATCATCGATGTGGAT 498
 QY 481 ATGCACATTTCTGTCATCATCTCTGCACTCATCTGTCACACTGCTGGGAGGG 540
 Db 499 ATGCACATTTCTGTCATCATCTCTGCACTCATCTGTCACACTGCTGGGAGGG 558
 QY 541 CTCATATCTGCGGCTACACATGATGTCGTTTCCAGCTCTTTTGCATTTTGTAGGGCTGTGG 600
 Db 559 CTCATATCTGCGGCTACACATGATGTCGTTTCCAGCTCTTTTGCATTTTGTAGGGCTGTGG 618
 QY 601 ATCAGCGTCCCTTTGCAATGTCACATCCCTGCACTGCGAGACATCGGGTTTCACTGCTGTG 660
 Db 619 ATCAGCGTCCCTTTGCAATGTCACATCCCTGCACTGCGAGACATCGGGTTTCACTGCTGTG 678
 QY 661 CATGCCAAATACCAAAAGCGTGGCTGGGAACCTGTTGACTCATCTGAAGTCTACTCTGTG 720
 Db 679 CATGCCAAATACCAAAAGCGTGGCTGGGAACCTGTTGACTCATCTGAAGTCTACTCTGTG 738
 QY 721 CTTGATAGTTTCTGTTGATGCTGGTGGGAATCCCATGGAAGCATACTTTTCAGAGG 780
 Db 739 CTTGATAGTTTCTGTTGATGCTGGTGGGAATCCCATGGAAGCATACTTTTCAGAGG 798
 QY 781 GTTCTCTCTTCTTCACGACCTATGCTCAAGTCTGCTCTTCTGTCAGCTTCGCG 840
 Db 799 GTTCTCTCTTCTTCACGACCTATGCTCAAGTCTGCTCTTCTGTCAGCTTCGCG 858
 QY 841 TGCCTGGTATGGCCATCCGACCATACTCATTTGGGGCATTTGGAGCATCAACAGACTGG 900
 Db 859 TGCCTGGTATGGCCATCCGACCATACTCATTTGGGGCATTTGGAGCATCAACAGACTGG 918
 QY 901 AACCAGACTGATATGGCTTCCAGATCCCAAGACTACAGAAGGAGGACATGATTTTA 960
 Db 919 AACCAGACTGATATGGCTTCCAGATCCCAAGACTACAGAAGGAGGACATGATTTTA 978
 QY 961 CCAATTTGCTGAGTATCTCTGCGCTGTGTATATTTCTTTTGTGCTTGGTGCAGTT 1020
 Db 979 CCAATTTGCTGAGTATCTCTGCGCTGTGTATATTTCTTTTGTGCTTGGTGCAGTT 1038
 QY 1021 TCTGCTGCTGTTATGTCATCAGCAGATTTCTTCCATCTTGTGAGCAAGTTCATGTTTGA 1080
 Db 1039 TCTGCTGCTGTTATGTCATCAGCAGATTTCTTCCATCTTGTGAGCAAGTTCATGTTTGA 1098
 QY 1081 CGAAGCATACACAGCTTTCTTTCAGACAAATGCTTCGACAAAGAAATCGTTTGGTT 1140
 Db 1099 CGAAGCATACACAGCTTTCTTTCAGACAAATGCTTCGACAAAGAAATCGTTTGGTT 1158
 QY 1141 ATCGCAATACACAGTGTGTTGTTGAGCATCTGCAACAGCCATGCGCTTGTGAGCAAA 1200
 Db 1159 ATCGCAATACACAGTGTGTTGTTGAGCATCTGCAACAGCCATGCGCTTGTGAGCAAA 1218
 QY 1201 ACTGTTATGGCTCTGTTACCTGAGTCTGACCTTGTGACCTGTTTACATGCTTATCTTCC 1260
 Db 1219 ACTGTTATGGCTCTGTTACCTGAGTCTGACCTTGTGACCTGTTTACATGCTTATCTTCC 1278
 QY 1261 CTGCTTGTGACCTCTGTTAAGGGAACCAACCTATGCGGCGGTGCGAGGTTATGTT 1320
 Db 1279 CTGCTTGTGACCTCTGTTAAGGGAACCAACCTATGCGGCGGTGCGAGGTTATGTT 1338
 QY 1371 TCTGCGCTTCTTCTGAGAATAACTGAGGAGGAGCCATATCTGATCTTCAGCCCTTGATC 1380
 Db 1339 TCTGCGCTTCTTCTGAGAATAACTGAGGAGGAGCCATATCTGATCTTCAGCCCTTGATC 1398
 QY 1381 TTTACCTCTGCTATACCTGATGATTAATGATATATAATCAGAAATTTCCATTTAAA 1440
 Db 1399 TTTACCTCTGCTATACCTGATGATTAATGATATATAATCAGAAATTTCCATTTAAA 1458
 QY 1441 ACATTGCCATGTTATCATCTTTAAACCAACATTTGCAATCTCTATCTAGCCAAAGTAT 1500
 Db 1459 ACATTGCCATGTTATCATCTTTAAACCAACATTTGCAATCTCTATCTAGCCAAAGTAT 1518
 QY 1501 CTATTTGAAAGTGAACCTTGCACCTAAATATAGATGATTTGATCTGCTGTTGCAAGA 1560
 Db 1519 CTATTTGAAAGTGAACCTTGCACCTAAATATAGATGATTTGATCTGCTGTTGCAAGA 1578

QY 1561 CACAGTGAAGAAAACATGATGAAGCAATTTCTGTCACAAAATGAAAATTAATTAATAGAT 1620
 Db 1579 CACAGTGAAGAAAACATGATGAAGCAATTTCTGTCACAAAATGAAAATTAATTAATAGAT 1638
 QY 1621 GAACCTTGACCTTGTGAAGCCACGACAGACATGACCCCTCAGCTCAACCTTTCACCAATAA 1680
 Db 1639 GAACCTTGACCTTGTGAAGCCACGACAGACATGACCCCTCAGCTCAACCTTTCACCAATAA 1698
 QY 1681 GAGCCCTTCCCTTGATGTTGATTCAGTCCAGAGGCTCTGGAGCTGAAGATAAATTACAG 1740
 Db 1699 GAGCCCTTCCCTTGATGTTGATTCAGTCCAGAGGCTCTGGAGCTGAAGATAAATTACAG 1758
 QY 1741 TGA 1743
 Db 1759 TGA 1761

RESULT 5
 AB043997
 LOCUS
 DEFINITION Homo sapiens mRNA for high-affinity choline transporter CHT1, complete cds.
 ACCESSION AB043997
 VERSION AB043997.1 GI:11231080
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 AUTHORS Okuda,T. and Haga,T.
 TITLE Functional characterization of the human high-affinity choline transporter
 JOURNAL FEBS Lett. 484 (2), 92-97 (2000)
 MEDLINE 20521663
 REFERENCE 2 (bases 1 to 5158)
 AUTHORS Okuda,T.
 TITLE Direct Submission
 JOURNAL Submitted (30-MAY-2000) Takashi Okuda, University of Tokyo, Faculty of Medicine, Department of Neurochemistry, 7-3-1 Hongo, Bunkyo-ku, Tokyo 1130033, Japan (E-mail:okuda@n.u-tokyo.ac.jp/neurochemistry, URL:http://park.ecc.u-tokyo.ac.jp/neurochemistry, Tel:81-3-5841-3560, Fax:81-3-6814-8154)
 FEATURES
 Location/Qualifiers
 1..5158
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="spinal cord"
 277..2019
 /codon_start=1
 /product="high-affinity choline transporter CHT1"
 /protein_id="BAB18161.1"
 /db_xref="GI:11231081"
 /translation="MAFHVEGLIAIVFYLLILVGIWAARTKNSGSAERSBAIV GGRDIGLVGFTMTATWGGYINGTAEEVYVPGYGLAWQAQDIGVLSLILGLFF AKPMRSKYVITMLDPFOQIYKRMGGLFLPALMGEMFAAIFSAIGATISVIDV MHISVTSIALIATLYLVGLYSVAYTDVQLGICFVGLWISVFLSHPAVADIGT AVHAKYQKPLWLTGVDSEVSWLDSFLLMLGGIPWQAYFORVLSSTSAVAYLSFL AAFGLVMAIPATILIGAGTADWNTQATTEADMLPIVQLCPVYISF FGUGSAVMSADSSILSSAFNNIYQLSFQNASDEIWMVNRITVVFVGSA TAMALITKYVGLWYLSLSDLYIVIPQLCLVLFVKGTNTYGAAGVSGVLFRTGG EPLYLQPLIFYPGYPDNDGNIYNQPKFLAMVTSFLTNICISVYLAKYLFESGILP PKLDVFDVAVRHSEENMDKTLVKNENIKLDELALVKPQSTLSSTFTTNKEAFLDV DSSPESGSGTDLQ"

BASE COUNT 1613 a 1019 c 981 g 1545 t
 ORIGIN
 Query Match 100.0%; Score 1743; DB 9; Length 5158;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGGCTTTCCATGTGGAGGACTGATAGTATCATCTGTTCTACCTTCTAATTTGCTG 60

DT	08-FEB-2002 (Rel. 70, Created)
DT	08-FEB-2002 (Rel. 70, Last updated, version 1)

Db 1561 CACAGTGAAGAGAACATGACAGACCACTTAGTCAGAAATGAAAACATCAAAATTAAT 1620
 QY 1621 GAACCTGCACTGTGAAGCCAGCAGACGATGACCTCAGCTCAACTTTCACCAATAA 1680
 Db 1621 GAACCTGCACTGTGAAGCCAGCAGACGATGACCTCAGCTCAACTTTCACCAATAA 1680
 QY 1681 GAGGCTTCCTTGATGTTGATTCAGTCCAGAGGCTCTGGGACTGAAGATAACTTACAA 1740
 Db 1681 GAGGCTTCCTTGATGTTGATTCAGTCCAGAGGATCTGGGACTGAAGATAACTTACAA 1740
 QY 1741 TGA 1743
 Db 1741 TGA 1743

RESULT 8
 AB030947
 LOCUS
 DEFINITION
 CCESSION
 ERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 FEATURES
 source

AB030947 4904 bp mRNA linear ROD 03-FEB-2000
 Rattus norvegicus mRNA for high-affinity choline transporter CHTL,
 complete cds.
 AB030947.1 GI:6863033
 choline transporter; high-affinity choline transporter CHTL.
 Rattus norvegicus (strain:Wistar) adult spinal cord cDNA to mRNA,
 clone_lib:rat spinal cord cDNA library clone:CHTL.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus
 1 (sites)
 Okuda,T., Haga,T., Kanai,Y., Endou,H., Ishihara,T. and Katsura,I.
 Identification and characterization of the high-affinity choline
 transporter
 Nat. Neurosci. 3 (2), 120-125 (2000)
 2 (bases 1 to 4904)
 Okuda,T.
 Direct Submission
 Submitted (09-AUG-1999) Takashi Okuda, University of Tokyo, Faculty
 of Medicine, Department of Neurochemistry; Hongo 7-3-1, Bunkyo-ku
 113-0033, Japan (E-mail:okuda@m.u-tokyo.ac.jp, Tel:+81-3-5841-3560,
 Fax:+81-3-3814-8154)
 Sequence updated (11-Jan-2000).
 Location/Qualifiers
 1. 4904
 /organism="Rattus norvegicus"
 /strain="Wistar"
 /db_xref="taxon:10116"
 /clone="CHTL"
 /tissue_type="spinal cord"
 /clone_lib="rat spinal cord cDNA library"
 /dev_stage="adult"
 224..1966
 /codon_start=1
 /product="high-affinity choline transporter CHTL"
 /protein_id="BAA90484.1"
 /db_xref="GI:6863034"
 /translation="MPFHVGLVAITLFLVGLIWAANKNGNAERSEAIIIV
 GGRDILVGGFTMTATWGGVINGTAENVYGGGLAWAQAPIGYSLILGLGF
 AKPMRSKGVYMLDPQQIYKRMGLLFLPALMFMFWAAAIIFSAIGATISVLDV
 VNISVTSALIALITLVGLYSVAYDVVOLFIFGLWISVPFALSHPAVDIGFT
 AVHAKYQSWLPALGIESVYTDNLNLLMLGIPWQAVFORVLSSTATYAOVLSFL
 APFGLMWPALFICIGAVTDNQTAYGFPDPKTEADMLIPVLOLCPVYISF
 FGLGAVSAAMVSADSIILSSAMFARNIYOLSFONASDKELVWVRIRYVFGASA
 TAMALLTKYVGLWYLSLVIIIFPOLLCVLFIKGTNYGAVAGYIFGLFRLITGG
 EPLYLQPLFVGYPPKNGLYNQRPFKTLSWTFNFTNICVSLAKYLFESGTP
 PKLIDFVAVSRHSENMEDKTLVRNENIKLDELAPVKPQSLTSLSTPTNEALIDV
 DSSPEGSGETDNLQ"

BASE COUNT 1447 a 991 c 939 g 1527 t
 ORIGIN

Query Match 80.0%; Score 1394.2; DB 10; Length 4904;
 Best Local similarity 87.5%; Pred. No. 0;

Matches 1525; Conservative 0; Mismatches 218; Indels 0; Gaps 0;
 QY 1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCTGCTTCTACCTCTTAATTTGGCTG 60
 Db 224 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCTGCTTCTACCTCTTAATTTGGCTG 60
 QY 61 GTTGAATATGGCTGCGCTGGAGAACCAAAACAGTGGCAGAGAGAGAGAGAGAGAGAGAG 120
 Db 284 GTTGAATATGGCTGCGCTGGAGAACCAAAACAGTGGCAGAGAGAGAGAGAGAGAGAG 120
 QY 121 GCCATCATAGTTGGTGGCCGAGATATGGTTTATTTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
 Db 344 GCCATCATAGTTGGTGGCCGAGACATTTGGTTTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
 QY 181 ACCTGGGTGCGGAGGAGGTATATCAATGGCAGACAGTGAAGCAGTGTATGTACAGGTTAT 240
 Db 404 ACCTGGGTGCGGAGGAGGTATATCAATGGCAGACAGTGAAGCAGTGTATGTACAGGTTAT 240
 QY 241 GGCTAGCTGGGCTCAGGACCAATTTGGTATTTCTTCTAGTCTGATTTTAGTGGGCGCTG 300
 Db 464 GGCTAGCTGGGCTCAGGACCAATTTGGTATTTCTTCTAGTCTGATTTTAGTGGGCGCTG 300
 QY 301 TTCTTTGCAAAACCTATGCGTTCAGAGGGGTATGTGACCATGTAGACCCCTTTTTCAGCAA 360
 Db 524 TTCTTTGCAAAACCTATGCGTTCAGAGGGGTATGTGACCATGTAGACCCCTTTTTCAGCAA 360
 QY 361 ATCTATGGAAGAGCATGGGCGGACCTCTCTTATTTCTGCTGACATGATGGGAGAAATGTT 420
 Db 584 ATCTATGGAAGAGCATGGGCGGACCTCTCTTATTTCTGCTGACATGATGGGAGAAATGTT 420
 QY 421 TGGGCTGCGACCAATTTCTCTGCTTTGGGAGCCACCATGAGTGTGATCATGATGATGAT 480
 Db 644 TGGGCTGCGACCAATTTCTCTGCTTTAGGGGTACCATGAGTGTGATCATGATGATGAT 480
 QY 481 ATGCACATTTCTGTCATCATCTGTCACATCTGTCACATCTGTCACACACTGGTGGGAGG 540
 Db 704 ATGCACATTTCTGTCATCATCTGTCACATCTGTCACATCTGTCACACACTGGTGGGAGG 540
 QY 541 CTCTATTCGTGGCTTACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
 Db 764 CTCTATTCGTGGCTTACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
 QY 601 ATCAGGCTCCCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 823
 Db 824 ATCAGGCTCCCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 823
 QY 661 CATGCAAAATACCAAAAGCCGTGGCTGGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 883
 Db 884 CATGCAAAATACCAAAAGCCGTGGCTGGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 883
 QY 721 CTGTAGTGTTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 943
 Db 944 CTGTAGTGTTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 943
 QY 781 GTTCTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1003
 Db 1004 GTTCTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1003
 QY 841 TGGCTGTGATGGCTTCCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
 Db 1064 TGGCTGTGATGGCTTCCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
 QY 901 AACCAAGTGTGATGGCTTCCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1123
 Db 1124 AACCAAGTGTGATGGCTTCCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1123
 QY 961 CCAATTTGCTGAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 Db 1184 CCAATTTGCTGAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 QY 1021 TCTG 1080
 Db 1244 TCTG 1080
 Db 1244 TCTG 1303

Tue Apr 1 13:57:23 2003

3 (bases 1 to 1743)
 Apparsundaram, S., Ferguson, S.M. and Blakely, R.D.
 Direct Submission
 Submitted (28-FEB-2001) Department of Pharmacology and Center for
 Molecular Neuroscience, Vanderbilt University, 23rd Avenue South at
 Pierce, Nashville, TN 37232-6420, USA
 Sequence update by submitter
 On Feb 28, 2001 this sequence version replaced gi:11527247.
 Location/Qualifiers
 1. .1743
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 1. .1743
 /codon_start=1
 /product="sodium and chloride-dependent high-affinity
 choline transporter"
 /protein_id="P36945.2"
 /db_xref="GI:13162670"
 /translation="MPFHEGLVAIILFYLLIFLGLIWAANKTKNSGNPERSEALIV
 GGRDILGVGFTMTATWGGYINGTAEAVYGGGLAWAHAPIGYSLILGLLEF
 AKPMRSKGYVMDLPEKQIYKRMGLLEFIPALMGEMFWAAIFSAIGATISVLDIGT
 VNLSIVSALIALTYLVGLYAVDVQVLFIFGLWISVPFALSHPAVDYDGL
 AVHAKTQSPWLGITIEVYVTLNDLFLMLGGIPQVAFQVLSVSSAYQAQVLSL
 AAFGLVMAIPALICIGASTDMWNOTAYCPDPKTEADMLPIVQLVCPVYISF
 FGLGAVSAAMSSADISILSSAFARNIYQLSPRONASDKELVWNRIVLVFGASA
 TAMALLTKYVGLWYSSDLYIIPOLLICVLFIKGTNTYNGAVAGYIFGLFLRTGS
 EPLYLQPLIFYPYSDKNGIYNQFPKTLNMTVFFNTICVSYLAKLYFESGTLF
 PKLDVDAVARHSEENMDKTLVNRNENIKLNELAPVKPRQSLTSLSTFTNKEALLDV
 DSPESGSGTEDNLQ"
 BASE COUNT 406 a 409 c 410 g 518 t
 ORIGIN
 Query Match 78.9%; Score 1375; DB 10; Length 1743;
 Best Local Similarity 86.8%; Pred. No. 0;
 Matches 1513; Conservative 0; Mismatches 230; Indels 0; Gaps 0;
 QY 1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCTGCTTCTACTCTCTAATTTTGGCTG 60
 DB 1 ATGGCTTTCCATGTGGAAGGACTGATAGCTATCATCTGCTTCTACTCTCTAATTTTGGCTG 60
 QY 61 GTTGAATATGGCTGCGCTGGAGAACCAAAACAGTGGCGAGAGAGAGAGAGAGAGAGAGAG 120
 DB 61 GTTGAATATGGCTGCGCTGGAGAACCAAAACAGTGGCGAGAGAGAGAGAGAGAGAGAGAG 120
 QY 121 GCATCATATAGTTGGTGGCGAGATATTTGTTTATGGTGGTGGATTTTACCATGACAGCT 180
 DB 121 GCATCATATAGTTGGTGGCGAGATATTTGTTTATGGTGGTGGATTTTACCATGACAGCT 180
 QY 181 ACCTGGTGGAGAGGGGTATATCATATGACAGCTGAAGCAGTGTATGTACCAGGTTAT 240
 DB 181 ACCTGGTGGAGAGGGGTATATCATATGACAGCTGAAGCAGTGTATGTACCAGGTTAT 240
 QY 241 GGCTAGCTTGGGCTCAGGACCAATTTGGATATTTCTTAGTCTGATTTTAGTGGGCTG 300
 DB 241 GGCTAGCTTGGGCTCAGGACCAATTTGGATATTTCTTAGTCTGATTTTAGTGGGCTG 300
 QY 301 TTTTTCGAAAACCTATGCTTCAAAAGGGTATGTGACCATGTAGACCCGTTTTCAGCAA 360
 DB 301 TTTTTCGAAAACCTATGCTTCAAAAGGGTATGTGACCATGTAGACCCGTTTTCAGCAA 360
 QY 361 ATCTATGGAACGCTATGGCGGAGCTCTCTTTTATTTCTGCTGCTGATGGAGAAATTTTC 420
 DB 361 ATCTATGGAACGCTATGGCGGAGCTCTCTTTTATTTCTGCTGCTGATGGAGAAATTTTC 420
 QY 421 TGGGCTGCAGCAATTTTCTGCTTTGGGAGCCACCATCATGAGTATCATGATGATGATGAT 480
 DB 421 TGGGCTGCAGCAATTTTCTGCTTTGGGAGCCACCATCATGAGTATCATGATGATGATGAT 480
 QY 481 GTGACATATTCGTGCTATCATCTCTGCACTCATCTGCCACTCTGTACACACTGCTGGGAGGG 540
 DB 481 GTGACATATTCGTGCTATCATCTCTGCACTCATCTGCCACTCTGTACACACTGCTGGGAGGG 540
 QY 541 CTCTATTCTGTGGGCTACAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REMARK
 COMMENT
 FEATURES
 source
 CDS
 1081 CGGAATATCTACAGCTTTCTTTCAGACAAATGCTCGGACAAAGAAATCGTTGGGTT 1140
 DB 1304 CGGAATATCTACAGCTTTCTTTCAGACAAATGCTCGGACAAAGAAATGCTGGGTC 1363
 QY 1141 ATGCGAATCAGAGT 1200
 DB 1364 ATGAGATCAGT 1423
 QY 1201 ACTGTGTATGGGCTGTGGTACCTGAGTCTGACCTTTGTGTGTGTGTGTGTGTGTGTGT 1260
 DB 1424 ACTGTGTATGGGCTGTGGTACCTGAGCTGTGACCTTTGTGTGTGTGTGTGTGTGTGT 1483
 QY 1261 CTGCTTTGT 1320
 DB 1484 CTGCTCTGT 1543
 QY 1321 TCTGGGCTCTTCTCGAATAAATGAGGGGAGCCATATCTGTATCTTTCAGCCCTTGATC 1380
 DB 1544 TTTGGACTTTTCTCGAATAAATGAGGGGAGCCATATCTATCTTTCAGCCCTTAATC 1603
 QY 1381 TTCTACCTGGCTATTACCTGATGATAATGATATATATATATATATATATATATATATAT 1440
 DB 1604 TTCTACCTGGCTATTACCTGATGATAATGATATATATATATATATATATATATATAT 1663
 QY 1441 ACATTGCCATGTTTACATCATTTCTTAACCAACATTTTGCATCTCTATCTAGCAAGTAT 1500
 DB 1664 ACATTGCCATGTTTACATCATTTCTTAACCAACATTTTGCATCTCTATCTAGCAAGTAT 1723
 QY 1501 CTATTTGAAAGTGGAACTTGGCCACCTAAATAGATGATTTGATGCTGTGTGTGCAAGA 1560
 DB 1724 CTATTTGAAAGTGGAACTTGGCCACCTAAATAGATGATTTGATGCTGTGTGTGCAAGA 1783
 QY 1561 CACAGTGAAGAAACATGATGATAAGCAATCTTCTGCAAAATGAAATATTAATATAGAT 1620
 DB 1784 CACAGTGAAGAAACATGATGATAAGCAATCTTCTGCAAAATGAAATATTAATATAGAT 1843
 QY 1621 GAATTCGACATGTGAAGCCACGACAGAGCATGACCCCTGACCTCACTTACCAATATAA 1680
 DB 1844 GAATTCGACATGTGAAGCCCTCGACAGAGCCTAACCTTCAGTTCACCAATATAA 1903
 QY 1681 GAGGCTTCTGTGATGTTGATTCAGTCCAGAGGCTCGAGAGGCTCGGACTGAAGATAATTA 1740
 DB 1904 GAGGCTTCTGTGATGTTGATTCAGTCCAGAGGCTCGAGAGGCTCGGACTGAAGATAATTA 1963
 QY 1741 TGA 1743
 DB 1964 TGA 1966
 RESULT 9
 AF276872
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1743)
 Apparsundaram, S., Ferguson, S.M. and Blakely, R.D.
 Molecular cloning and characterization of human and murine
 high-affinity choline transporters
 Unpublished
 2 (bases 1 to 1743)
 Apparsundaram, S., Ferguson, S.M. and Blakely, R.D.
 Direct Submission
 Submitted (09-JUN-2000) Department of Pharmacology and Center for
 Molecular Neuroscience, Vanderbilt University, 23rd Avenue South at
 Pierce, Nashville, TN 37232-6420, USA


```
Db 541 CTCTACTCTGTGGCATATACATGATGTTGTCCAGCTATTCTGCAATTTTATAGGACTGTGG 600
QY 601 ATCAGGCTCCCTTTGCAATGTCACATCTGCAGTGCAGAGACATCGGGTTCACTGCTGTG 660
Db 601 ATCAGTGTCCCTTTGCGCTGTGCATCTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGTG 660
QY 661 CATGCCAAATACAAAAGCCGTGGCTGGGAACCTGTTGACTCATCTCAAGTCTACTCTGTGG 720
Db 661 CATGCTAAATACCAAGAGTCCCTGGCTGGGAACCAATTGAATCAGTTGAAGTCTACACCTGG 720
QY 721 CTGTGATGTTTCTGTGTTGATGCTGGGTGGAATCCCAATGGCAACATCTTTCAGAGG 780
Db 721 CTGTGATGTTTCTGTGTTGATGCTGGGTGGAATCCCAATGGCAACATCTTTCAGAGG 780
QY 781 GTTCTCTCTTCTTCTCAGCCACCTATGCTCAAGTGTCTGCTCTCTCTGCTGCACTTTCGGG 840
Db 781 GTCTCTCTCTCTCTCAGCCACCTATGCTCAGTGTCTGCTCTCTCTGCTGCACTTTCGGG 840
QY 841 TGCTCTGGTATGSCCATCCACGCACTACTCATTTGGGGCCATTGGAGCATCAACAGACTGG 900
Db 841 TGCTCTGGTATGSCCTTACCCGCCATATGCTATGAGAGCTATTGGAGCTTCCACAGACTGG 900
QY 901 AACCAAGCTCATATGGGCTTCCAGATCCCAAGACTACAGAGAGCAGACATGATTTTA 960
Db 901 AACCAAGCTGCTACGGGTATCCAGATCCCAAGACTAAGGAGAGCAGACATGATTTCTC 960
QY 961 CCAATGTTCTGCAGTATCTCTGCGCTGTGTATATTTCTTTTGGTCTTGGTGCAGTT 1020
Db 961 CGATGCTGTGCAATGCTCTGCGCTGTGTATATTTCTTTTGGTCTTGGTGCAGTT 1020
QY 1021 TCTGCTGTGTATGTCATCAGCAGATCTTCCATCTTGTTCAGCAAGTTCCTGCTGAGTTCGA 1080
Db 1021 TCAGTGTGTGTCATGCTCTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
QY 1081 CGGACATCTACAGCTTCTTCTCAGACAAATGCTTGGGACAAAGAAATGCTTGGGTT 1140
Db 1081 CGGAAATCTACAGCTTCTTCTCAGACAAATGCTTGGGACAAAGAAATGCTTGGGTT 1140
QY 1141 ATGGAAATCAGAGTGTGTTGTTGGAGCATCTCAGACAGCAATGCTTCCGCCAG 1200
Db 1141 ATGAGGATCACTGTGTTGTTGGAGCATCTCAGACAGCAATGCTTCCGCCAG 1200
QY 1201 ACTGTGTATGGGCTCTGGTACCTCAGTCTGACCTTGTGTTTACATCGTATCTTCCGCCAG 1260
Db 1201 ACTGTGTATGGGCTCTGGTACCTCAGTCTGACCTTGTGTTTACATCGTATCTTCCGCCAG 1260
QY 1261 CTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGCGGTGGAGGTTATGTT 1320
Db 1261 CTGCTCTGTGTACTCTTCTCAGAGGAAACCAACCTATGGGCGGTGGAGGTTATGTT 1320
QY 1321 TCTGGCTCTTCTGAGAATAAATGGAGGGAGCCATATCTGTATCTTACGCTTGTATC 1380
Db 1321 TTTGGACTATCTCTGAGAATAAATGGAGGGAGCCATATCTGTATCTTACGCTTGTATC 1380
QY 1381 TTTACCTCTGCTTACCTGTATGATATGTTATGATATATAATCAGAAATTTCCATTTAA 1440
Db 1381 TTTACCTCTGCTTACCTGTATGATATGTTATGATATATAATCAGAAATTTCCATTTAA 1440
QY 1441 ACATTTGCCATGTTTACATCTTCTTAAACCAACATTTGCATCTCTCTATCTAGCAAGTAT 1500
Db 1441 ACTCTCTCCATGTTTACCTCTTCTTAAACCAACATTTGCATCTCTCTATCTAGCAAGTAT 1500
QY 1501 CTATTTGAAAGTGGAACTTGGCACTTAAATAGATGTTATTTGATGCTGTTGTGCAAGA 1560
Db 1501 CTATTTGAAAGTGGAACTTGGCACTTAAATAGATGTTATTTGATGCTGTTGTGCAAGA 1560
QY 1561 CACAGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
Db 1561 CACAGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
QY 1621 GAACCTGACCTTGTGAAGCCACGACAGAGCATGACCTCAGCTCAACTTTCACCAATAA 1680
Db 1621 GAACCTGACCTTGTGAAGCCACGACAGAGCATGACCTCAGCTCAACTTTCACCAATAA
```

```
Db 1621 GAACCTGACCTTGTGAAGCCACGACAGAGCTTAACCTCAGTTCAACTTTTCCACCAATAG 1680
QY 1681 GAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1740
Db 1681 GAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1740
QY 1741 TGA 1743
Db 1741 TGA 1743

RESULT 10
LOCUS BD012720
DEFINITION High-affinity choline transporter.
ACCESSION BD012720
VERSION BD012720.1 GI:22092909
KEYWORDS WO 0116315-A/4.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Haga,T. and Okuda,T.
High-affinity choline transporter
Patent: WO 0116315-A 4 08-MAR-2001;
JAPAN SCIENCE AND TECHNOLOGY CORP,TATSUYA HAGA,TAKASHI OKUDA
OS Mus musculus (mouse)
PN WO 0116315-A/4
PD 08-MAR-2001
PF 18-AUG-2000 WO 2000JP005545
PI 27-AUG-1999 JP 99P 240642,27-DEC-1999 JP 99P 368991
PC TATSUYA HAGA,TAKASHI OKUDA
PC C12N15/12,C07K14/47,C12Q1/68,C07K19/00,C07K16/18,C12N5/10, PC
A61K38/17,
PC A61K45/00,A61P25/28,G01N33/53,A01K67/027
CC
FH Key Location/Qualifiers
FT CDS (1)..(1743).
source /organism="Mus musculus"
BASE COUNT 407 a 410 c 409 g 517 t
ORIGIN

Query Match 78.8%; Score 1373.4; DB 6; Length 1743;
Best Local Similarity 86.7%; Pred. No. 0;
Matches 1512; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

QY 1 ATGGCTTTCCATGTTGGAGGACTGATAGCTATCATCTGTTCTACCTTCTTAATTTTCTG 60
Db 1 ATGCTTTTCCACGTAGAGGACTGGTAGCTATTATCTCTTCTACCTCTCTATATTTCTG 60
QY 61 GTTGAATATGGCTGCCTGGAGAACCAAAACAGTGGGAGCGAGCAAGAGCGAGCGAA 120
Db 61 GTTGAATATGGCTGCATGGAAACCAAAACAGCGGCAACCCAGAGAGACAGCTGAA 120
QY 121 GCCATCATAGTTGGTGGCCGAGATATTGGTTTATTTGGTGGTGGTATTTACCATGACAGCT 180
Db 121 GCCATCATAGTGGGCGGTGACATTTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
QY 181 ACCTGGGTGGAGGAGGGGTATATCAATGGCACAGCTCAAGCAGTTTATGTACCAAGTTAT 240
Db 181 ACCTGGGTGGAGGAGGGGTATACATCAATGGCACAGCAGCAGTGTATGGGCCAGGTTGT 240
QY 241 GGCTAGCTTTGGCTCAGCACCACCAATTGGATATTTCTTAGTCTGATTTTAGTGGGCTG 300
Db 241 GGCTAGCTTTGGCTCAGCACCACCAATTGGATATTTCTTAGTCTGATTTTAGTGGGCTG 300
QY 301 TTTCTTTGCAAAACCTATGCGTTTCAAGGGGTATGTGACCATGTTAGACCGGTTTTCAGCAA 360
Db 301 TTTCTTTGCAAAACCTATGCGTTTCAAGGGGTATGTGACCATGTTAGACCGGTTTTCAGCAA 360
```


Best Local Similarity 86.7%; Pred. No. 0;
Matches 1512; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

```
QY 1 ATGGCTTTCATGTGGAAGGACTGATAGCTATCATCGTGTCTACCTTCTAAATTTGCTG 60
DB 1 ATGCTTTCACGTACAGAGCTGGTGTATATCTCTTCTACCTCTTATATTTCTG 60
QY 61 GTTGAATATGGCTGCCTGGAGAACCAAAAACAGTGGCAGGCGCAGAGGCGACGAA 120
DB 61 GTTGAATATGGCTGCATGGGAAACCAAAAACAGCGGCAACCCAGAGAGGACAGTGA 120
QY 121 GCCATCATAGTGGTGGCCGAGATATGTTTATGTTGGTGGATTTACCATGACAGCT 180
DB 121 GCCATCATAGTGGGCGCGTACATGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
QY 181 ACCTGGTGGAGGAGGTATATCAATGGCACAGCTGAAGCACTTATGTACAGGTTAT 240
DB 181 ACCTGGTGGAGGAGGTATATCAATGGCACAGGAGGAGGAGGAGGAGGAGGAGGAG 240
QY 241 GGCTAGCTTGGGCTCAGGCACCAATTCGATATCTCTTGTAGTCTGATTTAGTGGCCTG 300
DB 241 GGCTAGCTTGGGCTCAGGCACCAATTCGATATCTCTTGTAGTCTGATTTAGTGGCCTG 300
QY 301 TTCTTTCGAAACCTATGCGTTCAAGGGTATGTGACCATGTTAGACCCGTTTCACAA 360
DB 301 TTTTTCGAAACCTATGCGTTTCGAGGGATATGTGACTATGTTAGACCCGTTTCACAA 360
QY 361 ATCTATGAAACGATGGGCGGACTCTCTGTTTATTCCTGACATGATGGGAGAAATGTC 420
DB 361 ATCTATGAAACGATGGGCGGACTCTCTGTTTATTCCTGACATGATGGGAGAAATGTC 420
QY 421 TGGGCTGACAGCAATTTCTCTGTTTGGGAGCCACCATCATGCGTGTATCTGATGTTG 480
DB 421 TGGGCTGACAGCAATTTCTCTGTTTGGGAGCCACCATCATGCGTGTATCTGATGTTG 480
QY 481 ATGACATATTTCTGCTATCATCTCTGACATCATCTGACATCTGACATCTGATGTTG 540
DB 481 GTGACATATTTCTGCTATCATCTCTGACATCATCTGACATCTGACATCTGATGTTG 540
QY 541 CTCTATCTGTCGCTTACATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 600
DB 541 CTCTATCTGTCGCTTACATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 600
QY 601 ATCAGCTGCTGCTTTCATGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
DB 601 ATCAGCTGCTGCTTTCATGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
QY 661 CATGCAATATCCAAAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
DB 661 CATGCAATATCCAAAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
QY 721 CTGATAGCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
DB 721 CTGATAGCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
QY 781 GTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
DB 781 GTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
QY 841 TGCCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 841 TGCCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
QY 901 AACGAGCTGATATGGCTTCCAGATCCCAAGACTACAGAGGAGGAGGAGGAGGAGGAGG 960
DB 901 AACGAGCTGCTTACGGTATCCAGATCCCAAGACTACAGAGGAGGAGGAGGAGGAGGAG 960
QY 961 CCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
DB 961 CCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
QY 1021 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
DB 1021 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
```

```
Db 1021 TCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
QY 1081 CGGAACATCTACAGCTTCTCTTCAGACAAATGCTTTCGACAAAGAAATCGTTGGGTT 1140
DB 1081 CGGAACATCTACAGCTTCTCTTCAGACAAATGCTTTCGACAAAGAAATCGTTGGGTT 1140
QY 1141 ATCGAATACAGCTGTTTGTGTTGGAGCATCTGCAACAGCCATGGCCTTGTGTCGAGAAA 1200
DB 1141 ATGAGGATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
QY 1201 ACTGCTGATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
DB 1201 ACTGCTGATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
QY 1261 CTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
DB 1261 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
QY 1321 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1321 TTTGGACTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
QY 1381 TTCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
DB 1381 TTCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
QY 1441 ACATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
DB 1441 ACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1501 CTATTTGAAAGTGGAACTTGGCCACTTAAATAGATGATGATGATGATGATGATGATGATG 1560
DB 1501 CTATTTGAAAGTGGAACTTGGCCACTTAAATAGATGATGATGATGATGATGATGATG 1560
QY 1561 CACAGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1620
DB 1561 CACAGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1620
QY 1621 GAATTTGCACTTGTGAAGCCACGACGATGATGATGATGATGATGATGATGATGATGATG 1680
DB 1621 GAATTTGCACTTGTGAAGCCACGACGATGATGATGATGATGATGATGATGATGATGATG 1680
QY 1681 GAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
DB 1681 GAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
QY 1741 TGA 1743
DB 1741 TGA 1743
```

RESULT 12
AX080443
LOCUS
DEFINITION
AX080443
ACCESSION
VERSION
AX080443.1
GI:13159872
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 4938)
Sierzege, M. and Albrandt, K.
Differentially expressed genes in the adipocytes of obese mice
Patent: WO 0078950-A 1 28-DEC-2000;
JOURNAL
AMYLIN PHARMACEUTICALS, INC. (US)
FEATURES
source
Location/Qualifiers
1..4938
/organism="Mus musculus"
/db_xref="taxon:10090"
/note="P4P6B1"
BASE COUNT
1436 a 1012 c 976 g 1514 t

ORIGIN									
Query Match 78.8%; Score 1373.4; DB 6; Length 4938;									
Best Local Similarity 86.7%; Pred. No. 0;									
Matches 1512; Conservative 0; Mismatches 231; Indels 0; Gaps 0;									
QY	1	ATGGCTTTCCATGTGGGAAGGACTGATAGTATCATCTGCTGTTTACCTTCTAATTTTGGTG	60						
DB	247	ATGTCTTTCCACGTAGAGGAGTGTAGTATATATCTCTTACCTCTTATCTTCTG	306						
QY	61	GTGGAATATGGCTGCTGGAGAACCAAAAACAGTGGCAGCGAGAGAGCGCAGCGAA	120						
DB	307	GTGGAATATGGCTGCTGGAGAACCAAAAACAGTGGCAGCGAGAGAGCGCAGCGAA	366						
QY	121	GCATCATATAGTTGGTGGCGGAGATATGGTTTATGGTTTGGTGGATTTACCATGACAGCT	180						
DB	367	GCATCATATAGTTGGTGGCGGAGATATGGTTTATGGTTTGGTGGATTTACCATGACAGCT	426						
QY	181	ACCTGGTGGAGGAGGTATATCAATGGCAGAGCTGAAGCAGTATATGTACCAAGTTAT	240						
DB	427	ACCTGGTGGAGGAGGTATATCAATGGCAGAGCTGAAGCAGTATATGTACCAAGTTAT	486						
QY	241	GGCTAGCTTGGCTCAGGCACCAATTTGGATATCTCTTACCTGATTTAGTGGCGCTG	300						
DB	487	GGCTAGCTTGGCTCAGGCACCAATTTGGATATCTCTTACCTGATTTAGTGGCGCTG	546						
QY	301	TTCTTTGCAAAACCTATGCTTCAAGGGGTATGTGACCATGTGTAGACCGCTTTACAGAA	360						
DB	547	TTCTTTGCAAAACCTATGCTTCAAGGGGTATGTGACCATGTGTAGACCGCTTTACAGAA	606						
QY	361	ATCTATGAAACATCGGCGGACTCCTGTTATCTCTGACCTGATGGGAGAAATGTC	420						
DB	607	ATCTATGAAACATCGGCGGACTCCTGTTATCTCTGACCTGATGGGAGAAATGTC	666						
QY	421	TGGGCTGAGCAATTTCTCTGCTTTGGGAGCCACCATGACGCTGATCGATCTGGAT	480						
DB	667	TGGGCTGAGCAATTTCTCTGCTTTGGGAGCCACCATGACGCTGATCGATCTGGAT	726						
QY	481	ATGCACATTTCTGTCATCATCTCTGCACTCATTTGCCATCTGTACACACTGGTGGGAGG	540						
DB	727	GTGACATATCGGTCATTTCTCTGCACTCATTTGCCATCTGTACACACTGGTGGGAGG	786						
QY	541	CTCTATTTCTGGGCTACACTGATCTCTGCACTCTTTTGGATTTTGTAGGGCTGTGG	600						
DB	787	CTCTATTTCTGGGCTACACTGATCTCTGCACTCTTTTGGATTTTGTAGGGCTGTGG	846						
QY	601	ATCAGGCTCCCTTGTGCAATTTGCATCTCTGCACTCTGTGAGAGCATCGGGTTCACTGCTG	660						
DB	847	ATCAGGCTCCCTTGTGCAATTTGCATCTCTGCACTCTGTGAGAGCATCGGGTTCACTGCTG	906						
QY	661	CATGCAAAATACAAAGCCGTGGCTGGGAGCTGTTGACTCATCTGAAGTCACTCTTGG	720						
DB	907	CATGCAAAATACAAAGCCGTGGCTGGGAGCTGTTGACTCATCTGAAGTCACTCTTGG	966						
QY	721	CTTGATAGTTTCTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTTCAGAGG	780						
DB	967	CTTGATAGTTTCTGTTGATGCTGGGTGGAATCCCATGGCAAGCATACTTTTCAGAGG	1026						
QY	781	GTCTCTCTCTTCTCTAGCCACTATGCTCAAGTCTGCTCTCTCTGGCAGCTTTCGGG	840						
DB	1027	GTCTCTCTCTTCTCTAGCCACTATGCTCAAGTCTGCTCTCTCTGGCAGCTTTCGGG	1086						
QY	841	TGCTGGTATGGCCATCCAGCACTACTCATTTGGGSCCATTTGGAGCATCAACAGACTGG	900						
DB	1087	TGCTGGTATGGCCATCCAGCACTACTCATTTGGGSCCATTTGGAGCATCAACAGACTGG	1146						
QY	901	AACAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAGAGGAGAGCATGATTTA	960						
DB	1147	AACAGACTGCATATGGGCTTCCAGATCCCAAGACTACAGAGAGGAGAGCATGATTTA	1206						
QY	961	CCAAATGTTCTGAGTATCTCTGCCCTGTGTATATTTCTTCTTCTTGTGCTGCAAT	1020						
DB	1207	CCGATCGTTCTGAGTATCTCTGCCCTGTGTATATTTCTTCTTCTTGTGCTGCAAT	1266						

QY	1021	TCCTGCTGTTATGTCATCAGCAGATTTCCATCTTGTCTCAGCAGATTTCCATGTTTGA	1080
DB	1267	TCAGTGTGTCATGTCCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1326
QY	1081	CGGAACATCTACAGCTTTTCTTCCACAAAATGCTTCGACAAAAGAAATTCGTTGGGT	1140
DB	1327	CGGAATATCTACAGCTTTTCTTCCACAAAATGCTTCGACAAAAGAAATTCGTTGGGT	1386
QY	1141	ATGCGAATCACAGTGTGTTGTTGAGAGATCTGCAACAGCCATGCTGCTGCTGCTGCTGCT	1200
DB	1387	ATGAGGATCACGTGCTGCTGTTGTTGAGAGATCTGCAACAGCCATGCTGCTGCTGCTGCT	1446
QY	1201	ACTGTGTATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1360
DB	1447	ACTGTGTATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1506
QY	1261	CTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1320
DB	1507	CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1566
QY	1321	TCGCGCTCTCTCTGAGAAATTAACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1380
DB	1567	TTTGGAGATATTTCTGAGAAATTAACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1626
QY	1381	TTCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1440
DB	1627	TTCTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1686
QY	1441	ACACTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1500
DB	1687	ACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1746
QY	1501	CTATTGAACTGGAACCTTGCACCTAAATAGATGATTTAGATGATTTAGATGATTTAGATGAT	1560
DB	1747	CTATTGAACTGGAACCTTGCACCTAAATAGATGATTTAGATGATTTAGATGATTTAGATGAT	1806
QY	1561	CACAGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1620
DB	1807	CACAGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1866
QY	1621	GAACTTGCACTTGTGAAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	1680
DB	1867	GAACTTGCACTTGTGAAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	1926
QY	1681	GAGGCTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1740
DB	1927	GAGGCTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1986
QY	1741	TGA 1743	
DB	1987	TGA 1989	
RESULT 13			
MU401467			
LOCUS			
DEFINITION			
Mus musculus mRNA for high affinity choline transporter (Chrl			
gene).			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			

1743 bp mRNA linear ROD 16-AUG-2000

MU401467

Mus musculus mRNA for high affinity choline transporter (Chrl gene).

ACCESSION AJ401467.1 GI:9843808

VERSION Chrl gene; high affinity choline transporter.

KEYWORDS house mouse.

SOURCE Mus musculus.

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 1743)

AUTHORS Wleand, A., Bonisch, H. and Bruss, M.

TITLE Molecular cloning of the human and murine high affinity choline transporters and characterization of the human gene-structure

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1743)

AUTHORS Brues, M.

1621	GAACCTTCAGCTTGTGAAGCCACGACAGCATGACCCCTCAGCTCAACTTTCCACCAATAAA	1680
1621	GAACCTTCAGCTTGTGAAGCCACGACAGCATGACCCCTCAGCTCAACTTTCCACCAATAAG	1680
1681	GAGCCCTTCCTTGATGTGATTCAGTCCAGGACGACGAGAGGCTGCGGACTCAAGATATTTACAG	1740
1681	GAGCCCTTCCTTGATGTGATTCAGTCCAGGACGACGAGAGGCTGCGGACTCAAGATATTTACAA	1740
1741	TGA	1743
1741	TGA	1743
RESULT 14		
TMA420808	2528 bp mRNA linear	VRT 27-NOV-2001
LOCUS	Torpedo marmorata mRNA for high affinity choline transporter (CHT1 gene).	
DEFINITION		
ACCESSION	AJ420808	GI:17148508
VERSION	CHT1 gene; high affinity choline transporter.	
KEYWORDS	marbled electric ray.	
SOURCE	Torpedo marmorata	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes; Elasmobranchii; Squalae; Hypnosqualea; Pristiogaster; Batoidae; Torpediniformes; Torpedinidae; Torpedo.	
REFERENCE	1. Guernonprez, L., O'Regan, S., Meunier, F.M. and Morot-Gaudry, F. J. Neurosci. 1998; 18: 1000-1010.	
AUTHORS		
TITLE	Cyclosporin, FK506 and rapamycin inhibit neuronal choline uptake via calcineurin-dependent and independent mechanisms	
JOURNAL	Unpublished	
REFERENCE	2. (bases 1 to 2528)	
AUTHORS	O'Regan, S.	
TITLE	Direct Submission	
JOURNAL	Submitted (21-NOV-2001) O'Regan S., Neurobiologie Cellulaire et Moléculaire, C.N.R.S., 1 av de la Terrasse, F-91198 Gif-sur-Yvette, FRANCE	
FEATURES	Location/Qualifiers	
source	1. .2528	
	/organism="Torpedo marmorata"	
	/db_xref="taxon:7788"	
	/clone="th312"	
	/tissue_type="electric lobe"	
	/tissue_lib="lambda ZAPII ELL"	
gene	49. .2492	
	/gene="CHT1"	
CDS	49. .1803	
	/function="neuronal Na-dependent choline transporter"	
	/codon_start=1	
	/evidence=experimental	
	/product="high affinity choline transporter"	
	/protein_id="CAD12727.1"	
	/db_xref="GI:17148509"	
	/translation="MTVHIDGIVAVLFLVLLFVGLWAANKSKNTSMGAMDRSEAI	
	MTGGDGLLVGGFTMTATWGGVINGTAEAVYVPGYGLWAQAQPFYALSLVIGGL	
	FFAKPMSRGVYMLDPQOMYKRGMLFPALLGEIWFSAILSALGATLSVID	
	ININVSVAIVANIVLYGLYSVAVDVVQLFCIFGLWISIPFALLNPATVDII	
	VTANQVQEPWVGNIGSLIINDELLMLGIPQVYFORVLSASATYAOVLS	
	FLAFCVCLMAIPSLVIAIGTSTWNTSYGLPGIPKNETDMILPIVLQHLCPPI	
	SFGLGAAVSAAMSSADSLSSASSFARNIYHLAFQBSADKEIYVWNRITFLFGG	
	AATSMALLAQIPYVPCYLDHSGFDVYVQRPFPKTMALFSLFGNGVSVLYKLF	
	GGPVLHMQPFIYVPCYLDHSGFDVYVQRPFPKTMALFSLFGNGVSVLYKLF	
	VSGILPPLKDLFSLVSVSKHSKIMDKTFLMNQDNITLSLVHVNPHSASVSAALTNK	
	EAFEDIEPNPELSKSDN"	
	2487. .2492	
	/gene="CHT1"	
polya_signal	674 a 496 g 525 g 833 t	
BASE COUNT	674 a 496 g 525 g 833 t	
ORIGIN		
Query Match	49.78; Score 867; DB 5; Length 2528;	

QY 1075 TTGCGACGGAACATCTACACGCTTCTTCCAGACAAATGCTTCGGACAAAGAAATCGTT 1134
 Db 1129 TTGCTCGGAATATTTACCATCTTCTTTCAGACAGAGCTTCAGACAAAGAAATAGTG 1188
 QY 1135 TGGGTTATGCGAATCAGAGTGTGTTGTTGGAGCATCTGCAACAGCCATGGCCTGCTG 1194
 Db 1189 TGGGTAATGCGAATCAGCATATTTCTATTTGGAGGAGCTGCAACATCTATGGCATTGCTT 1248
 QY 1195 AGGAAACTGTGTATGGGCTCTGGTACCTCAGTTCAGCTTCTGACCTTGTACATCGTTATCTTC 1254
 Db 1249 GCTCAATCAATCTAGGCTCTGGATCTGAGCTCAGATCTTGTCTACGTCATTATCTTT 1308
 QY 1255 CCCCAGCTGCTTTGTGTACTCTTTGTTAAGGGAACCAACACCTATGGGGCCGTTGGCAGGT 1314
 Db 1309 CCTCAATTAATATCAGTGTCTTCGTCAGAGGGAACCAACACATATGGTCTATGCTGA 1368
 QY 1315 TATGTTTCTGGGCTCTTCTCGAGATACTGGAGGAGGAGCCATATCTGTATCTTCAGCCC 1374
 Db 1369 TATATCATTTGGCTTTTCTCGGATTTAGTGGTGGTGAACCATATTTACATATGCAAGCCA 1428
 QY 1375 TTGATCTCTACCTGGCT-----ATTACCTGATGATAATGATATATATAAT 1422
 Db 1429 TTTATTTATACCTGGATGCTATTAGATCATCTCTTTGGAGATGATCTGTTTATGTT 1488
 QY 1423 CAGAAATTTCCATTTAAACACTTCCCATGTTTACATCATCTTAAACCAACATTTGCATC 1482
 Db 1489 CAGAGATTTCCCTTTAAACCAATGGAATGTTTATCTCTTGGGCAACACATGCTGTA 1548
 QY 1483 TCCTATCTAGCAAGTATCTATTTGAAAGTGGAACTTGGCCACCTAAATAGATGATTT 1542
 Db 1549 TCATATCTGTCAAGTACCTGTTCTGAAGTGGAAATTTGCCACCAAAATAGACCTCTT 1608
 QY 1543 GATGCTGTGTTGCAAGACAGACAGTCAAGAAACATGGAATGGAATCTTGTCAAAAT 1602
 Db 1609 GACAGCTGTGATCAAAACAGTAAGAAATCATGGAACAAACATCTTGAATGATCAG 1668
 QY 1603 GAAATATTAATAGATGAACTTGCACCTTGTGAAGCCACAGACAGATGACCTCAGC 1662
 Db 1669 GACAATATTTCTTGTGAGAGCTGTGTCATCTTAATCAATACAGATGCTTCATTTAGT 1728
 QY 1663 TCACTTTCAACAAAGAGGCTTCTTGTGATGTTGATTCAGTCCAGAGGCTCTGGG 1722
 Db 1729 GCTGCTTGACCAATAGGAAGCATTTGAAGACATTTGAGCAATCTGCAATCTTCTAAG 1788
 QY 1723 ACTGAAGATAATTTA 1737
 Db 1789 TCAGGCAATGATTGA 1803
 RESULT 15
 HSA308384
 LOCUS
 DEFINITION Homo sapiens partial CHT1 gene for high affinity choline transporter, exon 9.
 ACCESSION AJ308384
 VERSION AJ308384.1 GI:18369775
 KEYWORDS ChT1 gene; high affinity choline transporter.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Wieland, A., Bonisch, H. and Bruss, M.
 TITLE Molecular cloning of the human and murine high affinity choline transporters and characterization of the human gene structure
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2239)
 AUTHORS Bruss, M.
 TITLE Direct Submission
 JOURNAL Submitted (23-JAN-2001) Bruess M., University of Bonn, Pharmacology and Toxicology, Reuterstr. 2b, D-53113 Bonn, GERMANY
 FEATURES Location/Qualifiers

source 1..2239
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2q11-13"
 1..2239
 /gene="CHT1"
 <1..791
 /gene="CHT1"
 /number=8
 792..>2239
 /gene="CHT1"
 /number=9
 /usedin=AJ308378:CHT1_CDS
 /usedin=AJ308377:CHT1_mRNA
 BASE COUNT 664 a 442 c 440 g 693 t
 ORIGIN
 Query Match 36.2%; Score 630.8; DB 9; Length 2239;
 Best Local Similarity 99.7%; Pred. No. 2.4e-163;
 Matches 632; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1110 AAATGCTTCGGACAAAGAAATCGTTGGGTTATGCGAATCAGAGTGTGTTGTTGGAGC 1169
 Db 788 ACAGGCTTCGGACAAAGAAATCGTTGGGTTATGCGAATCAGAGTGTGTTGTTGGAGC 847
 QY 1170 ATCTGCAACAGCCATGGCTTGTGTCAGAAACTGTGTATGGGCTCTGCTACCTCAGTTC 1229
 Db 848 ATCTGCAACAGCCATGGCTTGTGTCAGAAACTGTGTATGGGCTCTGCTACCTCAGTTC 907
 QY 1230 TGACCTTGTGTACATCATCTTATCTTCCCCAGCTGCTTGTGTACTCTTGTGTTAAGGAA 1289
 Db 908 TGACCTTGTGTACATCATCTTATCTTCCCCAGCTGCTTGTGTACTCTTGTGTTAAGGAA 967
 QY 1290 CAACACCTATGGGCGCTGGCAGGTTATGTTTTCGGGCTCTTCTCGAATAAATGAGG 1349
 Db 968 CAACACCTATGGGCGCTGGCAGGTTATGTTTTCGGGCTCTTCTCGAATAAATGAGG 1027
 QY 1350 GGAGCCATATCTGATCTTCTCAGCCCTTGATCTTCTACCTGGCTATTACCTGATGATA 1409
 Db 1028 GGAGCCATATCTGATCTTCTCAGCCCTTGATCTTCTACCTGGCTATTACCTGATGATA 1087
 QY 1410 TGGTATATATATCAGAAATTTCCATTTAAACACTTGGCCATGTTTACATCATCTTAAAC 1469
 Db 1088 TGGTATATATATCAGAAATTTCCATTTAAACACTTGGCCATGTTTACATCATCTTAAAC 1147
 QY 1470 CAACATTTGCATCTCTCTATCTAGCCCAAGTATCTATTGAAAGTGGAACTTGGCCACTAA 1529
 Db 1148 CAACATTTGCATCTCTCTATCTAGCCCAAGTATCTATTGAAAGTGGAACTTGGCCACTAA 1207
 QY 1530 ATTAGATGATTGATGCTGTTGTCAGACACAGTGAAGAAACATGGATAAGACAAT 1589
 Db 1208 ATTAGATGATTGATGCTGTTGTCAGACACAGTGAAGAAACATGGATAAGACAAT 1267
 QY 1590 TCTTGTCAAAAATGAAAATATTAAATAGATGAACCTTGCACCTTGTGAAGCCACAGAG 1649
 Db 1268 TCTTGTCAAAAATGAAAATATTAAATAGATGAACCTTGCACCTTGTGAAGCCACAGAG 1327
 QY 1650 CATGACCTCAGCTCAACTTTCCACCAATAAAGAGGCTTCTCTTGATGTTGATTCAGTCC 1709
 Db 1328 CATGACCTCAGCTCAACTTTCCACCAATAAAGAGGCTTCTCTTGATGTTGATTCAGTCC 1387
 QY 1710 AGAAGGCTCTGGGACTGAGATAAATTTACAGTGA 1743
 Db 1388 AGAAGGCTCTGGGACTGAGATAAATTTACAGTGA 1421

Search completed: March 31, 2003, 17:49:36
 Job time : 4580 secs

Tue Apr 1 13:57:25 2003

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 16:34:01 ; Search time 67 Seconds
(without alignments)

1153.513 Million cell updates/sec

US-10-069-541-6

Perfect score: 2972

Sequence: 1 MAFHVEGLIAIVFVLLILL.....EAFLDVDSPEGSGTEDNLQ 580

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2972	100.0	580	22 AAB74665	Human high affinity
2	2972	100.0	580	22 AAB86837	Human CHOT protein
3	2820	94.9	580	22 AAB74664	Rat high affinity
4	2810	94.5	580	22 AAB72388	Mouse P4P61 OWA (
5	2801	94.2	580	22 AAB74666	Mouse high affinity
6	1453	48.9	576	22 AAB74663	C. elegans high af
7	723.5	24.3	278	22 AAB71601	Drosophila melanog
8	344	11.6	492	22 AAB96805	Putative P. abyssal
9	311.5	10.5	675	22 AAB85102	Novel human transp
10	310.5	10.4	675	22 AAB60093	Human transport pr

11	310.5	10.4	675	23 ABB80599	Human sbg1007026SG
12	308.5	10.4	662	16 AAR73595	Cotransporter prot
13	306	10.3	660	22 ABG30151	Novel human diago
14	298.5	10.0	675	23 AAU77134	Human sodium-sugar
15	298	10.0	672	16 AAR73593	Cotransporter prot
16	287	9.7	659	20 AAY31221	Human SAAT1 protei
17	283	9.7	659	20 AAY31221	Putative P. abyssal
18	277.5	9.3	596	22 AAE08088	Human transporter-
19	277.5	9.3	596	22 AAE08088	Human protein havi
20	277.5	9.3	596	23 ABB80588	Human sbg1020829SG
21	277.5	9.3	596	23 AAO14199	Human transporter
22	272.5	9.2	524	22 AAG93195	C glutamicum prote
23	272.5	9.2	524	22 AAB76757	Corynebacterium gl
24	272.5	9.2	718	22 ABG01897	Novel human diago
25	272.5	9.2	718	23 AAB47976	BCW2. Homo sapien
26	272.5	9.2	718	23 AAG78575	Human SLC5A3 amino
27	271	9.1	537	22 AAB96626	Putative P. abyssal
28	262.5	8.8	518	23 ABP39899	Staphylococcus epi
29	260	8.7	681	23 AAO14202	Human transporter
30	260	8.7	752	22 ABG28100	Novel human diago
31	260	8.7	752	23 AAE16778	Human transporter
32	248	8.3	742	23 AAE16778	Chlamydia pneumoni
33	232.5	7.8	465	20 AAY35106	Human NIS. Homo s
34	231	7.8	643	22 AAB73920	Human protein NOV5
35	230.5	7.8	664	23 AAU85407	Drosophila melanog
36	222	7.5	587	22 ABB62630	Novel human protei
37	222	7.5	591	23 ABP61007	C glutamicum prote
38	219	7.4	551	22 AAG90677	Corynebacterium gl
39	219	7.4	551	22 AAB76764	Rat sodium/iodide
40	218	7.3	618	18 AAW32105	Rat NIS. Rattus s
41	218	7.3	618	22 AAB73921	Human protein havi
42	216.5	7.3	618	22 AAE06584	Human transporter
43	216.5	7.3	618	23 AAO14195	Human membrane tra
44	209	7.0	635	21 AAY71060	Staphylococcus epi
45	206.5	6.9	530	23 ABP39665	

ALIGNMENTS

RESULT 1
AAB74665
ID AAB74665 standard; Protein; 580 AA.
XX
AC AAB74665;
XX
XX 01-JUN-2001 (first entry)
DT
DE Human high affinity choline transporter protein.
DE High affinity choline transporter; cho-1; Alzheimer's disease;
KW diagnosis.
KW Homo sapiens.
OS
PN WO200116315-A1.
XX
PD 08-MAR-2001.
PF
PF 18-AUG-2000; 2000WO-JP05545.
XX
XX 27-AUG-1999; 99JP-0240642.
PR 27-DEC-1999; 99JP-0368991.
XX
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
PA
PA Haga T, Okuda T;
XX
XX WPI: 2001-226688/23.
DR N-ESDB; AAF81712.
XX
XX New rat and human spinal cord high affinity choline transporters,
XX useful in diagnosis of Alzheimer's disease and screening promoters as

PT drugs for treating Alzheimer's disease -
 PS Claim 8; Page 76-78; 90pp; Japanese.
 CC The present sequence represents a human (Homo sapiens) high affinity
 CC choline transporter protein designated cho-1. The cho-1 protein has
 CC norepinephrine and neuroprotective activities. The cho-1 polynucleotide and
 CC protein can be used for the diagnosis of diseases related to the
 CC expression of cho-1 by comparing the cho-1 polynucleotide sequence in a
 CC sample to that of a control. Drug compositions containing the cho-1
 CC protein or expression promoters or inhibitors of cho-1 are useful for
 CC treating disorders characterized by abnormal levels of cho-1, such as
 CC Alzheimer's disease.

SQ Sequence 580 AA;
 Query Match 100.0%; Score 2972; DB 22; Length 580;
 Best Local Similarity 100.0%; Pred. No. 2.3e-289;
 Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAFHVEGLIAIIVFYLLILVGIWAARTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
 DB 1 MAFHVEGLIAIIVFYLLILVGIWAARTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60
 QY 61 TWGGGGYINGTAAVYVPGYGLAWAQAIPIGYSLILGGLFFAKPMRSKGYVTMLDPPQQ 120
 DB 61 TWGGGGYINGTAAVYVPGYGLAWAQAIPIGYSLILGGLFFAKPMRSKGYVTMLDPPQQ 120
 QY 121 IYKRMGGLLFIIPALMGEMFWAAAFSALGATISVIIDVDMHISVILSALITATLYLVGG 180
 DB 121 IYKRMGGLLFIIPALMGEMFWAAAFSALGATISVIIDVDMHISVILSALITATLYLVGG 180
 QY 181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGTAVHAKYQKPMWLTVDSEYISW 240
 DB 181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGTAVHAKYQKPMWLTVDSEYISW 240
 QY 241 LDSFLLMLGGIPQWQAYFQVLSSTATYAQVLSFLAAGCLVMAIPAILIGASTDW 300
 DB 241 LDSFLLMLGGIPQWQAYFQVLSSTATYAQVLSFLAAGCLVMAIPAILIGASTDW 300
 QY 301 NOTAYGLPDPKTEADMLPIVQLCPVYISFFGLGAVSAVMSADSSILSASSMFA 360
 DB 301 NOTAYGLPDPKTEADMLPIVQLCPVYISFFGLGAVSAVMSADSSILSASSMFA 360
 QY 361 RNIYQLSFRQNASDKREIVVMRITVFVFGASATAMALLTKTVGLWYLSDDLVIYIFPQ 420
 DB 361 RNIYQLSFRQNASDKREIVVMRITVFVFGASATAMALLTKTVGLWYLSDDLVIYIFPQ 420
 QY 421 LLCVLFVKGTNTYGAAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480
 DB 421 LLCVLFVKGTNTYGAAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480
 QY 481 TLAMVTSFLTNCISYLAKEFLFESGTLPPKLDVFDVAVARSEENMDKTLVKNENIKLD 540
 DB 481 TLAMVTSFLTNCISYLAKEFLFESGTLPPKLDVFDVAVARSEENMDKTLVKNENIKLD 540
 QY 541 ELALVKPRQSMSTLSTTNKFAFLDVSSPESGSGTEDNLQ 580
 DB 541 ELALVKPRQSMSTLSTTNKFAFLDVSSPESGSGTEDNLQ 580

RESULT 2
 AAB86837
 ID AAB86837 standard; Protein; 580 AA.
 XX
 AC AAB86837;
 XX
 DT 26-NOV-2001 (first entry)
 XX
 DE Human CHOT protein.
 XX
 KW CHOT; human; choline transporter; chromosome 2q11-13; norepinephrine;
 KW neuroprotective; gene therapy; antisense therapy; degenerative disease;

cognitive disorder; Alzheimer's disease.

Homo sapiens.

DE10009055-Al.

30-AUG-2001.

28-FEB-2000; 2000DE-1009055.

28-FEB-2000; 2000DE-1009055.

(BRUE/) BRUESS M.

(BOEN/) BOENISCH H.

Brueess M, Boenisch H;

WPI: 2001-590709/67.

N-PSDB; AAH49207.

A new gene encoding human choline transporter, designated hCHOT is located on chromosome 2q11-13 and is useful to treat degenerative disorders such as Alzheimer's disease.

Disclosure; Page 11; 12pp; German.

This invention describes a novel gene encoding human choline transporter, designated hCHOT which is located on chromosome 2q11-13. The products of the invention have norepinephrine and neuroprotective activity and can be used for gene or antisense therapy. (I) is used to treat degenerative disease, particularly cognitive disorders such as Alzheimer's disease. Sense and antisense oligonucleotides derived from the gene may be used in diagnostics and other techniques. This sequence represents the human CHOT protein described in the invention.

SQ Sequence 580 AA;

Query Match 100.0%; Score 2972; DB 22; Length 580;

Best Local Similarity 100.0%; Pred. No. 2.3e-289;

Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFHVEGLIAIIVFYLLILVGIWAARTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

DB 1 MAFHVEGLIAIIVFYLLILVGIWAARTKNSGSAEERSEAIIVGGRDIGLLVGGFTMTA 60

QY 61 TWGGGGYINGTAAVYVPGYGLAWAQAIPIGYSLILGGLFFAKPMRSKGYVTMLDPPQQ 120

DB 61 TWGGGGYINGTAAVYVPGYGLAWAQAIPIGYSLILGGLFFAKPMRSKGYVTMLDPPQQ 120

QY 121 IYKRMGGLLFIIPALMGEMFWAAAFSALGATISVIIDVDMHISVILSALITATLYLVGG 180

DB 121 IYKRMGGLLFIIPALMGEMFWAAAFSALGATISVIIDVDMHISVILSALITATLYLVGG 180

QY 181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGTAVHAKYQKPMWLTVDSEYISW 240

DB 181 LYSVAYTDVVQLFCIFVGLWISVPFALSHPAVADIGTAVHAKYQKPMWLTVDSEYISW 240

QY 241 LDSFLLMLGGIPQWQAYFQVLSSTATYAQVLSFLAAGCLVMAIPAILIGASTDW 300

DB 241 LDSFLLMLGGIPQWQAYFQVLSSTATYAQVLSFLAAGCLVMAIPAILIGASTDW 300

QY 301 NOTAYGLPDPKTEADMLPIVQLCPVYISFFGLGAVSAVMSADSSILSASSMFA 360

DB 301 NOTAYGLPDPKTEADMLPIVQLCPVYISFFGLGAVSAVMSADSSILSASSMFA 360

QY 361 RNIYQLSFRQNASDKREIVVMRITVFVFGASATAMALLTKTVGLWYLSDDLVIYIFPQ 420

DB 361 RNIYQLSFRQNASDKREIVVMRITVFVFGASATAMALLTKTVGLWYLSDDLVIYIFPQ 420

QY 421 LLCVLFVKGTNTYGAAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480

DB 421 LLCVLFVKGTNTYGAAGYVSGFLRITGGEPYLYLQPLIFYPGYPPDDNGIYNQKFPFK 480

Tue Apr 1 13:57:25 2003

QY 481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVDAVVAHSEENMDKTIIVKNNIKLD 540
 DB 481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVDAVVAHSEENMDKTIIVKNNIKLD 540
 QY 541 ELALVKPRQSMSTLSTFTNKEAFLVDSSPEGSGETDNLQ 580
 DB 541 ELALVKPRQSMSTLSTFTNKEAFLVDSSPEGSGETDNLQ 580

RESULT 3
 AAB74664
 ID AAB74664 standard; Protein; 580 AA.
 AC AAB74664;
 DT 01-JUN-2001 (first entry)
 DE Rat high affinity choline transporter protein.
 KW High affinity choline transporter; cho-1; Alzheimer's disease;
 KW diagnosis.
 OS Rattus norvegicus.

XX WO200116315-A1.
 PN 08-MAR-2001.
 PD 18-AUG-2000; 2000WO-JP05545.
 PF 27-AUG-1999; 99JP-0240642.
 PR 27-DEC-1999; 99JP-0368991.
 XX (NTSC-) JAPAN SCI & TECHNOLOGY CORP.
 PA Haga T, Okuda T;
 PI WPI; 2001-226688/23.
 DR N-PSDB; AAF81711.
 XX New rat and human spinal cord high affinity choline transporters,
 PT useful in diagnosis of Alzheimer's disease and screening promoters as
 PT drugs for treating Alzheimer's disease
 XX Claim 5; Page 69-71; 90pp; Japanese.

XX The present sequence represents a rat (Rattus norvegicus) high affinity
 CC choline transporter protein designated cho-1. The cho-1 protein has
 CC nootropic and neuroprotective activities. The cho-1 polynucleotide and
 CC protein can be used for the diagnosis of diseases related to the
 CC expression of cho-1 by comparing the cho-1 polynucleotide sequence in a
 CC sample to that of a control. Drug compositions containing the cho-1
 CC protein or expression promoters or inhibitors of cho-1 are useful for
 CC treating disorders characterised by abnormal levels of cho-1, such as
 CC Alzheimer's disease.
 XX Sequence 580 AA;

Query Match 94.9%; Score 2820; DB 22; Length 580;
 Best Local Similarity 93.1%; Pred. No. 4.3e-274;
 Matches 540; Conservative 24; Mismatches 16; Indels 0; Gaps 0;

QY 1 MAFHVEGLIAIVFYLLILLVGINAAWTKNSGSAERSEAIIVGGRIIGLVGGFTMTA 60
 DB 1 MPFHVEGLVAIFLLFYLLVGINAAWTKNSGSAERSEAIIVGGRIIGLVGGFTMTA 60
 QY 61 TWVGGYINGTAEAYVPGYGLAWAQAIPGYSLSLILGGLFFAKPMRSKGYVTMLDPQQ 120
 DB 61 TWVGGYINGTAEAYVPGYGLAWAQAIPGYSLSLILGGLFFAKPMRSKGYVTMLDPQQ 120
 QY 121 IYGRMGGLLFTPALMGEMFWAAAFPSALGATISVIIDVDMHISVIATLTLVGG 180
 DB 121 IYGRMGGLLFTPALMGEMFWAAAFPSALGATISVIIDVDMHISVIATLTLVGG 180

QY 181 LYSVAYTDVVQLFCIFVGLWISVPPFALSHPAVDIGTAVHAKYQKPLGTVDSEVYSW 240
 DB 181 LYSVAYTDVVQLFCIFVGLWISVPPFALSHPAVDIGTAVHAKYQKPLGTVDSEVYSW 240
 QY 241 LDSFILLMLGGIPQWAFQVRLSSSATYAQVLSFSLAAGCLVMAIPAILIGAIGASTDW 300
 DB 241 LDNFMILLMLGGIPQWAFQVRLSSSATYAQVLSFSLAAGCLVMAIPAILIGAIGASTDW 300
 QY 301 NOTAYGLDPDKTTEADMLPIVLQYLCVPIYISFFGGLGAVSAAMSSADSSILSASSMFA 360
 DB 301 NOTAYGPPDKTTEADMLPIVLQYLCVPIYISFFGGLGAVSAAMSSADSSILSASSMFA 360
 QY 361 RNIYQLSFRONASDKIIVWMRIIVFVGASATAMALITKTVYGLWLSDDLVIYIPPPQ 420
 DB 361 RNIYQLSFRONASDKIIVWMRIIVFVGASATAMALITKTVYGLWLSDDLVIYIPPPQ 420
 QY 421 LLCVLFGKNTYGAAGYVSGFLRITGGPEYLYLQPLIFPGYYPDNGIYNOKPFPK 480
 DB 421 LLCVLFGKNTYGAAGYVSGFLRITGGPEYLYLQPLIFPGYYPDNGIYNOKPFPK 480
 QY 481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVDAVVAHSEENMDKTIIVKNNIKLD 540
 DB 481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVDAVVAHSEENMDKTIIVKNNIKLD 540
 QY 541 ELALVKPRQSMSTLSTFTNKEAFLVDSSPEGSGETDNLQ 580
 DB 541 ELALVKPRQSMSTLSTFTNKEAFLVDSSPEGSGETDNLQ 580

RESULT 4
 AAY72388
 ID AAY72388 standard; Protein; 580 AA.
 XX AAY72388;
 AC AAY72388;
 DT 24-APR-2001 (first entry)
 DE Mouse P4P6B1 OMA (obese mice adipocyte) protein.
 KW Mouse; OMA protein; obese mice adipocyte; P4P6B1;
 KW fuel metabolism disorder; therapy; obesity; diabetes; gene therapy;
 KW anorectic; antidiabetic.
 XX Mus sp.
 XX WO200078950-A2.
 PN 28-DEC-2000.
 PD 13-JUN-2000; 2000WO-US16217.
 PF 22-JUN-1999; 99US-0141515.
 PR (AMYL-) AMYLIN PHARM INC.
 PA Sierzega M, Albrandt K;
 PI WPI; 2001-112322/12.
 DR N-PSDB; AAD02457.
 XX Novel obese mice adipocyte polypeptides useful in diagnosis and
 XX treatment of disorders of fuel metabolism such as obesity or diabetes
 XX Claim 11; Fig 3; 83pp; English.

XX The present sequence is mouse OMA (obese mice adipocyte) protein
 CC encoded by P4P6B1 cDNA. The P4P6B1 cDNA fragment was generated by
 CC RNA fingerprinting using random primers P4 and P6.
 CC OMA is used as a diagnostic reagent for diagnosing a disorder of fuel
 CC metabolism in an underweight or an overweight individual, by detecting
 CC the transcription level of a gene encoding OMA, which is induced or
 CC

repressed in an individual by a factor such as genetic obesity, fasting and refeeding of a fasted individual. OMA is useful in the generation of antibodies, for use in pharmaceutical compositions and for studying DNA/protein interactions. Nucleic acids encoding OMA are involved in gene therapy. An inhibitor of OMA or an antisense oligonucleotide that inhibits expression of OMA are useful for treating disorders of fuel metabolism such as obesity or diabetes.

XX Sequence 580 AA;

Query Match 94.5%; Score 2810; DB 22; Length 580;
Best Local Similarity 93.1%; Pred. No. 4.3e-273;
Matches 540; Conservative 23; Mismatches 17; Indels 0; Gaps 0;

QY 1 MAFHVEGLIAIVFYLLLLGIWAARTKNSGSAERSEAIIVGGDGLLVGGFTMTA 60
Db 1 MSFHVGEGLVAIILFYLLLLGIWAARTKNSGSAERSEAIIVGGDGLLVGGFTMTA 60
QY 61 TWVGGYINGTAEAVYVPGYGLAWAQAIPYGLSLIILGGLFFAKPMRSKGYVTMLDPFQ 120
Db 61 TWVGGYINGTAEAVYVPGYGLAWAQAIPYGLSLIILGGLFFAKPMRSKGYVTMLDPFQ 120
QY 121 IYGRMGGLLIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALITATLYLVGG 180
Db 121 IYGRMGGLLIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALITATLYLVGG 180
QY 181 LYSVAYTDVQVLCIFVGLWISVFPALSHPAVDIGTAVHAKYQKPLGTVDSSSEVSW 240
Db 181 LYSVAYTDVQVLCIFVGLWISVFPALSHPAVDIGTAVHAKYQKPLGTVDSSSEVSW 240
QY 241 LDSFLLMLGGIPQWQAYFQVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAISTDW 300
Db 241 LDSFLLMLGGIPQWQAYFQVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAISTDW 300
QY 301 NOTAYGLPDKTTEADMILPIVQLCPVYISFFGLGAVSAAMSSADSSILSASSMFA 360
Db 301 NOTAYGLPDKTTEADMILPIVQLCPVYISFFGLGAVSAAMSSADSSILSASSMFA 360
QY 361 RNIYQLSFRQNASDKIEIWMRTIVFVGASATAMALLTKTVYGLWLSDDLVIYIFPQ 420
Db 361 RNIYQLSFRQNASDKIEIWMRTIVFVGASATAMALLTKTVYGLWLSDDLVIYIFPQ 420
QY 421 LLCVLFVKGTNTYGVAGYVSGLEFRITGGEPLYLQPLIFYPGYDDNGIYNQKFPK 480
Db 421 LLCVLFVKGTNTYGVAGYVSGLEFRITGGEPLYLQPLIFYPGYDDNGIYNQKFPK 480
QY 481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVARHSEENMDKTLVKNENIKLD 540
Db 481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVARHSEENMDKTLVKNENIKLD 540
QY 541 ELALVKPQSMTSLSTFTNKEAFLDVDSPEGSGTEDIQLQ 580
Db 541 ELALVKPQSMTSLSTFTNKEAFLDVDSPEGSGTEDIQLQ 580

RESULT 5

AAB74666

ID AAB74666 standard; Protein; 580 AA.

XX AC AAB74666;

XX DT 01-JUN-2001 (first entry)

XX DE Mouse high affinity choline transporter protein.

XX KW High affinity choline transporter; cho-1; Alzheimer's disease;

XX OS diagnosis.

XX OS Mus musculus.

XX OS WO200116315-A1.

XX OS 08-MAR-2001.

XX PD

XX 18-AUG-2000; 2000WO-JP05545.
XX 27-AUG-1999; 99JP-0240642.
XX 27-DEC-1999; 99JP-0368991.
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX Haga T, Okuda T;

XX WPI; 2001-226688/23.

XX N-PSDB; AAF81713.

XX New rat and human spinal cord high affinity choline transporters, useful in diagnosis of Alzheimer's disease and screening promoters as drugs for treating Alzheimer's disease

XX Claim 11; Page 82-85; 90pp; Japanese.

XX The present sequence represents a mouse (Mus musculus) high affinity choline transporter protein designated cho-1. The cho-1 protein has neurotropic and neuroprotective activities. The cho-1 polynucleotide and protein can be used for the diagnosis of diseases related to the expression of cho-1 by comparing the cho-1 polynucleotide sequence in a sample to that of a control. Drug compositions containing the cho-1 protein or expression promoters or inhibitors of cho-1 are useful for treating disorders characterised by abnormal levels of cho-1, such as Alzheimer's disease.

XX Sequence 580 AA;

Query Match 94.2%; Score 2801; DB 22; Length 580;

Best Local Similarity 92.8%; Pred. No. 3.5e-272;

Matches 538; Conservative 23; Mismatches 19; Indels 0; Gaps 0;

QY 1 MAFHVEGLIAIVFYLLLLGIWAARTKNSGSAERSEAIIVGGDGLLVGGFTMTA 60
Db 1 MSFHVGEGLVAIILFYLLLLGIWAARTKNSGSAERSEAIIVGGDGLLVGGFTMTA 60
QY 61 TWVGGYINGTAEAVYVPGYGLAWAQAIPYGLSLIILGGLFFAKPMRSKGYVTMLDPFQ 120
Db 61 TWVGGYINGTAEAVYVPGYGLAWAQAIPYGLSLIILGGLFFAKPMRSKGYVTMLDPFQ 120
QY 121 IYGRMGGLLIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALITATLYLVGG 180
Db 121 IYGRMGGLLIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALITATLYLVGG 180
QY 181 LYSVAYTDVQVLCIFVGLWISVFPALSHPAVDIGTAVHAKYQKPLGTVDSSSEVSW 240
Db 181 LYSVAYTDVQVLCIFVGLWISVFPALSHPAVDIGTAVHAKYQKPLGTVDSSSEVSW 240
QY 241 LDSFLLMLGGIPQWQAYFQVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAISTDW 300
Db 241 LDSFLLMLGGIPQWQAYFQVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAISTDW 300
QY 301 NOTAYGLPDKTTEADMILPIVQLCPVYISFFGLGAVSAAMSSADSSILSASSMFA 360
Db 301 NOTAYGLPDKTTEADMILPIVQLCPVYISFFGLGAVSAAMSSADSSILSASSMFA 360
QY 361 RNIYQLSFRQNASDKIEIWMRTIVFVGASATAMALLTKTVYGLWLSDDLVIYIFPQ 420
Db 361 RNIYQLSFRQNASDKIEIWMRTIVFVGASATAMALLTKTVYGLWLSDDLVIYIFPQ 420
QY 421 LLCVLFVKGTNTYGVAGYVSGLEFRITGGEPLYLQPLIFYPGYDDNGIYNQKFPK 480
Db 421 LLCVLFVKGTNTYGVAGYVSGLEFRITGGEPLYLQPLIFYPGYDDNGIYNQKFPK 480
QY 481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVARHSEENMDKTLVKNENIKLD 540
Db 481 TLAMVTSFLTNICISYLAKYLFESGTLPPKLDVFDVAVARHSEENMDKTLVKNENIKLD 540
QY 541 ELALVKPQSMTSLSTFTNKEAFLDVDSPEGSGTEDIQLQ 580
Db 541 ELALVKPQSMTSLSTFTNKEAFLDVDSPEGSGTEDIQLQ 580

Matches	132;	Conservative	99;	Mismatches	196;	Indels	118;	Gaps	25;
QY	8	LIAIIIVFYLLILVLGWAARWTKNSGAERSEAILVGGRDIGLLVGGFTMTATWGGGY	67						
Db	14	LVAFLFTLILPILGVFAMKRKS-----EEDFFVGGRAMDKITVAASVSSGRSSL	66						
QY	68	INSTAEAVVPVPGYLAWAQAPTGYSLU-----LILGLFFAKPMRSGKYVTMLDPFOOIYG	123						
Db	67	VLGLSGMAYKMGTVAVW---AAVGYIVAEMFQFVYMGIIRLKFSERNATVPDYFEARFR	124						
QY	124	K-----RMGG-----LLFI PALGEMFWAAAIFSA LGATISVIIDVDHSHIISALIALTL	174						
Db	125	DTSKILRIARAIIIIIIFLTSYVGAOFNAG-----KTLSTALGISIFTALMISVLMII V	178						
QY	175	YLVGGLYSVAVTDVQLFCIFVGLWISVPPFALSHPAVADIGFT-----AVHAKYOK	226						
Db	179	YMLGGFTA VAYNDVRIVMI GLW-----LPVIAVAKVGGTEBVLKV LHALDPKLIN	233						
QY	227	PW----LGTVDSESYSWLDSFLLLMLG-GIPWOAY-FORVLSSSSATYA QVLSFLAAFGC	281						
Db	234	PWAFGAGVVG-----FLGIGFGSPQPHIIVRYMSIDDPNKLRVSTVVGTFWN	282						
QY	282	LVMAIPA I LTGAIGASTDWNQTAYCLDPDKTT--EADMI LP-IVLOYLC PYYISFFGIGL	338						
Db	283	VVLAWGAIFVGLAGRAI-----VPDVSQLPKNAEMIYPYLSAQYFPPI LVGIL-IG	333						
QY	339	AVSAAVNSSDSTSILSASSMFARNIYQLSFQNA--SDKEIVWMYMRITVFVFGSATAMA	396						
Db	334	GIFAAILSTADSQLVVVASTVVKDLQYEVKKIKIDEKTALTISRVTIVLVGFALA ILA	393						
QY	397	LRTKTYVGYLWLSDDLVI-VIVF-----POLLCVLFVKGTNTY GAVAGYVSGLFL	445						
Db	394	-----YVAKDIIFWFVLEFAWGGLGASFGLTLLSLYWKGTTKWGLAGMIVGIT	443						
QY	446	RITGEPYLYXLOPLIYPGYPDNDGIYNQKPFKTLAMVTSFLTNICISYLA KYLPESG	505						
Db	444	TIW---KYLKPI-----TGly-ELVP-----AFIFS LIATIIVSMITK-----	479						
QY	506	TLPPK 510							
Db	480	--PPE 482							
RESULT	9								
AAB85102									
ID	AAB85102	standard; Protein; 675 AA.							
AC	AAB85102;								
XX									
DT	22-AUG-2001	(first entry)							
DE		Novel human transporter protein (NHP).							
XX									
KW		Novel human protein; NHP; sodium; glucose; co-transporter; therapeutic;							
KW		diagnostic; pharmacogenomic; gene therapy; mental disorder.							
OS		Homo sapiens.							
XX									
PN	WO200142469-A1.								
XX									
PD	14-JUN-2001.								
XX									
PF	08-DEC-2000; 2000WO-US33363.								
XX									
PR	10-DEC-1999; 99US-0170137.								
XX		(LEXI-) LEXICON GENETICS INC.							
PI	Doncho G, Scoville J, Turner CA, Friedrich G, Zambrowicz B;								
PI	Abuin A, Sands AT;								
XX									
DR	WPI; 2001-381689/40.								
DR	N-Psdb; AAF84022, AAF84023.								

Tue Apr 1 13:57:25 2003

XX New nucleic acid molecules encoding human proteins that are
PT structurally similar to sodium-glucose co-transporters, useful for
PT therapeutic, diagnostic, and pharmacogenomic applications
XX
XX Claim 4; Page 29-30; 34pp; English.
XX
XX The invention relates to a novel human protein (NHP) that is structurally
CC similar to sodium-glucose co-transporters. The NHP protein and DNA are
CC useful in therapeutic, diagnostic and pharmacogenomic applications. An
CC oligopeptide comprising at least 12 amino acids of the NHP is useful for
CC diagnosis, drug screening, clinical trial monitoring and the treatment of
CC physiological disorders. The NHP nucleic acids are useful for identifying
CC mutations associated with a particular disease, and also in a diagnostic
CC for producing NHP, and in gene therapy. NHP polypeptides, its fragments
CC or mutants are useful for generating antibodies, as reagents in
CC diagnostic assays, for the identification of other cellular gene products
CC related to NHP, and as reagents in assays for screening compounds that
CC are useful as pharmaceutical reagents in the treatment of mental,
CC biological or medical disorders and diseases. The present sequence
CC represents the amino acid sequence of the human NHP.
XX
XX Sequence 675 AA;
XX
XX Query Match 10.5%; Score 311.5; DB 22; Length 675;
XX Best Local Similarity 23.0%; Pred. No. 4.9e-22;
XX Matches 152; Conservative 112; Mismatches 235; Indels 161; Gaps 28;
XX
XX 2 AFHVEGL---IATIVFY-LLLLVGIWAARWTKNGSAERSEAIIVGGRDGLLVGGF 56
XX 18 AFPOKLEPGDIAVLVFLFLVAVGLNSTVTK-----RDTVKGYFLAGGDMVMPVGA 72
XX 57 TMTATWVGCGYINGTAEEAVVPGYGLAWAQAIPGYSLI-LGGLF-----FAKPMRS 108
XX 73 SLFASNVGSGHF-----ICLAGSSAATGISVSAYELNGLFSVLMWLAWIFLPIYI 121
XX 109 KGYTMDLPPQIYKGRMGLLFPALMGEMFWAAAFSAL-----GAT-IVSIIDVDM 161
XX 122 AGQVTMPPEYLR---KREGIR-IPILLAVLYFIYIKTSVDMYAGAIQOSSLHDL 177
XX 162 HSIWISALIAFLYTLVGLSVAYTDVQLFCIVGLWISVPALSHPAVADIGTAVH 221
XX 178 YLAIVGELLATAVYTVAGGLAAVIYTDALQTLMLIGALTLMGY--SFAAVG--GMEGLK 233
XX 222 AKY-----QKPMIGTVDSSEVYS-WLDSFLLML 249
XX 234 ERYFLALASNRSENSSCGLPREDAFHIFRDLTSLDLPWPGVLFMGSIPLWY----- 285
XX 250 GGIPW---QAYFORVLSSSAYIAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQATY 306
XX 286 ---WCTDQIVQRTLAANKLSHAKGALAAVLYKVLPLFIMVPGMWSKILFPDQVA-- 339
XX 307 LPDPKTEE-----ADMILPILVQLCPVYISFPGLCAGSAAYMSSADSSILSASM 358
XX 340 CADEICQKICNSPGSCSDIAPKLVLELLPTGLRGLMVAVMVAALMSLTSIFNSASHI 399
XX 359 FARNIYOLSFQNASDKIEIVWVRITVFVFGASATAMALLTKVYIGLW----- 406
XX 400 FTMDLWN-HLRPRASEKELMTVGRVF-----LLLVLSILWIPVWQASGGOL 447
XX 407 ---YLLSSDLVYI-----VIFPOLLCVLFVKGNTYGVAGVYVSGFLRITG-GEPLYLQ 458
XX 448 FYIQISYISLQPPVAVF---INGCFWKRTNEKGAFWGLISGLLLGLVRLVLDVFIYQ 504
XX 459 LIIFYPGYPPDDNGIYNQRPFTLAMVTSFLTNICISYLAKEYFESGTLPLKLDV----- 513
XX 505 RC-----DQDERPVLKSIHLYFSMILSTVLTIVSYVWF-----TEPSPKEMVSHLT 555
XX 514 ---FDVAVARHSEENMDKTIIVKNENIKLD-----ELALVKPRQSMTSLSTFNKEA 562
XX 556 WFTRHDPVVOKEQAPPAAPLSLTLSONGMPAESSSSSVQFEMVOENTSKTSCDMTPKQS 615

RESULT 10
AAB60093 standard; Protein; 675 AA.
XX
XX AAB60093;
XX 28-MAR-2001 (first entry)
XX Human transport protein TPPT-13.
XX Human; transport protein; TPPT; transport disorder; metabolic disorder;
XX neurological disorder; cardiovascular disorder; reproductive disorder;
XX immune disorder; cancer.
XX Homo sapiens.
XX WO200078953-A2.
XX 28-DEC-2000.
XX 16-JUN-2000; 2000WO-US16668.
XX 17-JUN-1999; 99US-0139923.
XX 10-AUG-1999; 99US-0148177.
XX 18-AUG-1999; 99US-0149357.
XX 28-OCT-1999; 99US-0162287.
XX (INCY-) INCYTE GENOMICS INC.
XX Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;
XX Baughn MR, Asimzai Y, Lu DAM, Au-Young J, Patterson C;
XX WPI: 2001-041424/05.
XX N-PSDB; AAF27713.
XX Isolated polypeptide with a human transport protein sequence is useful
XX for the diagnosis, prevention and treatment of disorders associated
XX with the immune, reproductive and cardiovascular systems -
XX Claim 2; Page 115-116; 165pp; English.
XX The present invention provides the protein and coding sequences for 43
XX novel human transport proteins (designated TPPTs). These can be used in
XX the diagnosis and treatment of transport, metabolic, neurological,
XX reproductive, cardiovascular and immune disorders, and cell proliferative
XX disorders such as cancer.
XX
XX Sequence 675 AA;
XX
XX Query Match 10.4%; Score 310.5; DB 22; Length 675;
XX Best Local Similarity 23.0%; Pred. No. 6.2e-22;
XX Matches 151; Conservative 114; Mismatches 237; Indels 155; Gaps 28;
XX
XX 2 AFHVEGL---IATIVFY-LLLLVGIWAARWTKNGSAERSEAIIVGGRDGLLVGGF 56
XX 18 AFPOKLEPGDIAVLVFLFLVAVGLNSTVTK-----RDTVKGYFLAGGDMVMPVGA 72
XX 57 TMTATWVGCGYINGTAEEAVVPGYGLAWAQAIPGYSLI-----LILGGLFAKPMRSKY 111
XX 73 SLFASNVGSGHFGLA-----GSGAATGISVSAYELNGLFSVLMWLAWIFL--PIVIAQ 124
XX 112 VTMLDFPQIYKGRMGLLFPALMGEMFWAAAFSAL-----GAT-IVSIIDVDMHIS 164
XX 125 VTTMPEYLR---KREGIR-IPILLAVLYFIYIKTSVDMYAGAIQOSSLHDLILA 180
XX 165 VIISALITATLYLVGLSVAYTDVQLFCIVGLWISVPALSHPAVADIGTAVHAKY 224
XX 181 IVGLLATAVYTVAGGLAAVIYTDALQTLMLIGALTLMGY--SFAAVG--GMEGLKEY 236
XX 225 -----QKPMIGTVDSSEVYS-WLDSFLLMLGGI 252
XX 237 FLALASNRSENSSCGLPREDAFHIFRDLTSLDLPWPGVLFMGSIPLWY----- 285

us-10-069-541-6.rag

Tue Apr 1 13:57:25 2003

QY 437 AGYVSGFLRI-----TG-----GEPYLYLOPLIFVPGYPPDNGIY 473
 Db 487 WGLVLGLIGISRMITEFAYGTGSCMEPSNCTIICGVHLYFAILF-----534
 QY 474 NQKFPKTLAMVTSFLTNICISYLAFLFESGTLPLPKLDFDVAVA-RUSENMOKTILV 532
 Db 535 -----VSIITVVVSLFTKPI-----PDVHLYRLCWSLRNKE-----568
 QY 533 KNEIKID--ELALVKPROSMTLSSTFTNKEAF-----LDVDSPEGSSTED 577
 Db 569 --ERIDLAGEEDIOEAPEATDEVPKKKKFFRRAYDLFCGLDQDKGPKMTKEE 623

RESULT 13
 ABG30151
 ID ABG30151 standard; Protein; 560 AA.

AC ABG30151;
 XX 18-FEB-2002 (first entry)
 DT Novel human diagnostic protein #30142.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 PR (HYSE-) HYSEQ INC.

Query Match 10.4%; Score 308.5; DB 16; Length 562;
 Best Local Similarity 23.4%; Pred. No. 9.6e-22;
 Matches 154; Conservative 110; Mismatches 238; Indels 155; Gaps 26;

QY 11 IIVFYLLLVIGIWAARTKNSGSAERSEALIVGGRDGLLVGGFTMTATWVGGYING 70
 Db 32 IIVFYLLLVIGIWAARTKNSGSAERSEALIVGGRDGLLVGGFTMTATWVGGYING 86
 QY 71 TAEAVVPCYGLAWAQAIPIGYS-----LSLILGLLFFAKPMRSKGYVTMLDPFQIY-GK 124
 Db 87 LA-----GTGAASGIATGTFEWNALIMVVLGVVFPYIIRA-GVVTMPYLQRRFGK 139
 QY 125 RMGGLLFPALMGEMFW--AAIFSAIGAT-LSVILIDVDMHISVIISALIALTYLTVGL 181
 Db 140 RIQYLSLSLLYIFTKISADIFS--GALFQLTGLDIYVAIILLVITGLYTITGL 197
 QY 182 YSVAYTDVQLFCIFVGLMISVPFALSHPAVADIGTAVHAKY-----Q 225
 Db 198 AAVIYDTLTQATMVGSVILTLGGIPW-----QAYFQVLSSSSA 267
 QY 226 KPWLTGTVDSSEVYSWLSDFLLMLGGIPW-----QAYFQVLSSSSA 307
 Db 254 KCYTPREDAFHI-----FRDAITGDPWPGVLFGVMSILTLWTCIDQIVQCLSAKN 321
 QY 268 TYAQLVSLAFLGCVLMAIPAILGATGASTDMNQTAYGLPDP-----KTTEADMLP 321
 Db 308 SHVRAGCTLCGLYLVMPFLLVMMGWSRILYTKVACVVPSECEYCGRVGCTNTAF 367
 QY 322 IVLOQLCPVYISFFGLGAVSAVMSADSSILSSAFNIIYOLSFQNASDKIEIYVWM 381
 Db 368 TLVVELMPLNGRLGLMSVMAASLSLTSIFNSASTLTMDIY-TKIRKASEKELMAG 426
 QY 382 RI-TVVFVGASTAMALLTKTVYG--LWYLSLSDLYVI--VIFPQLLCVLFVKGNTYGA 436
 Db 427 RLFMLFLIGISTAWPIVQSAQSGOLFQDYIQSITSYLGPPIAAVFLLAIFWKRVPNEGAF 486

Claim 20; SEQ ID No 60510; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving a food supplement. (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, for genetic disorders or other traits to assess biodiversity responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AGC0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

[illegible]

us-10-069-541-6.rag

Tue Apr 1 13:57:25 2003

```

403 DLWN-HURPRASEKELMIVGRFV-----LLLVLSILWIPVQASOGQLFIY 450
408 LSSDLVYI-----VIFPQLLCVLFVKGNTYGVAGVYVSGFLRLITG-GEPLYILOPLIF 461
451 IQSISYLQPPVAVVF-----IMGEWKRNEKGAFWGLISGLLGLVRLVLDIFYVQPRC- 506
462 YPGYPPDNGIYNQKFPKTKAMTSLTNICISYLAFLFESGTLPPKLDV----- 513
507 ---DQDERPLVLSIHLYFSMILSTVLITVTSVNF-----TEPSKEMVSHLTWFT 558
514 -FDVAVARHSENMDDTKTLVKNENIKLD-----ELALVPRQSMTSSFTNKEA 562
559 RHPVVOREQAPPAPLSLITLSONGMPAEASSSSVQFEMVOENTSKTHSCDMTPKQS 615

RESULT 15
AAR73593
ID AAR73593 standard; Protein; 672 AA.
XX AAR73593;
XX 26-DEC-1995 (first entry)
XX Cotransporter protein SNST1.
XX Sodium ion nucleoside cotransporter protein; SNST1; rabbit.
XX Oryctolagus cuniculus.
XX US5410031-A.
XX 25-APR-1995.
XX 24-FEB-1992; 92US-0841651.
XX 24-FEB-1992; 92US-0841651.
XX (REGC ) UNIV CALIFORNIA OFFICE TECHNOLOGY.
XX Pajor AM, Wright EM;
XX WPI; 1995-169636/22.
XX N-PSDB; AAQ89779.
XX Mammalian sodium ion nucleoside cotransporter protein (SNST) cDNA -
XX used to produce SNST for identifying nucleoside transport inhibitors
XX and e.g. nucleoside analogues capable of effective uptake into cells
XX
XX Disclosure; Columns 17-22; 25pp; English.
XX
XX cDNAs were isolated from a rabbit kidney library by high stringency
XX hybridization with a rabbit sodium/glucose cotransporter (SGLT)
XX cDNA. Clone RK-C was demonstrated by expression studies to encode
XX most of an SNST protein. A composite sequence (AAQ89779) was obt'd. by
XX combining RK-C and another clone, RK-44. The cDNA was used to
XX express SNST1 in Xenopus oocytes.
XX
XX Sequence 672 AA;
XX
XX Query Match 10.0%; Score 298; DB 16; Length 672;
XX Best Local Similarity 25.0%; Pred. No. 1.1e-20;
XX Matches 153; Conservative 89; Mismatches 232; Indels 138; Gaps 25;
XX
XX 9 IAIT-VFYLLILLGWAARTKNSGAERSEATVGGRIIGLVGGFTWTATVWGGY 67
XX 26 IAVIAAYELLVIGVGLWSMCRN-NGTV---GGYFLAGSNVWMPVGLASFASNIGSGH 80
XX 68 INGTRAEAVYVPGYGLAWAQAPIGYLSL---LILGLLFFAKPMRSKGVYVTLDPFQIYG 123
XX 81 FVGLA-----GTGAANGLAAGFEWNNALFVVLGLLGNLFPAPVYLITAGVITM-----PQYL 130
XX 124 KRMGG-----LLFTPALMGEMFWAAAF--SALGATISVIIDVDMHISVISA 169

```

```

131 KRFGGHRIRLYSLVLSFLYIFTKISVDMFSGAVFIQALGNWI-----YASVIAL 182
170 LIATLYTLVGGLYSVAYTDVQVQLFCIFVGLWISVPFALSHPAVADIGTAVHAKY----- 224
183 GITVYTYTGGLAALMYDTQTFVILAGAFILTCYAFHEVG-----GYSGLFDKYMAMT 238
225 -----OKFWLGHVDSSEVYSWLDSELLL---MLGGIPW-----QAYF 258
239 SLTVSEDAVGNISSCYRPRDPSYHLRDPVTGDLFWPALLLGLLTIVSGWYWCSDQVIV 298
259 QVLSSSSATYAQVLSFLAAFGCLYMAIPAILIGAISTDWNQTAIGLPDPKT-----TE 314
299 QRCLAGRNLTHTKAGCILGKYLKTPMFLWMPGMISRLYPDEVACVAPEVCKRVCGTE 358
315 E-ADMILPVLIOYLCPPVVISFEGLGAVSAANWSSADSSILSASSMFARNIYOLSFRQA 372
359 VGCSNIAYPRLVVKLMPNGLRGLMLAVMLAALMSSLRSIFNSSLTFTMDIYTL--RPRA 416
373 SDKEIVWVRITVFEVFGASATAMALTKTVG---LWYLSDDLVIYIV--IFPQLLCVLFV 427
417 GEGELLVGRLMWVVFIVAVSVAVLPPVQAAQGGQLFDYIQSVSSYLAPPVSAVVFVALFV 476
428 KGTNTYGAVAGVYVSGFLRLITGGEPLYILOPLIFYPGYPPDDNGIYNQKFPKTKAMV-- 485
477 PRVNEKGAFWGLIGLLMGLARLP-----EFSFGTGCVRP 513
486 ---TSLTNICISYLAFLFESG-----TLP-PKLDVFDVAVVA-RHSENMDDKI 530
514 SACPALCRVHYLYFAIVLFFCSGLLIIVSLCTAPIPRKHLHLVFLSRHSE----- 567
531 LVKNENIKLDEL 542
568 --EREDLDADEL 577

Search completed: March 31, 2003, 18:34:31
Job time : 70 secs

```



```
RT transporter.";  
RL FEBS Lett. 484:92-97(2000).  
RP [5]  
RA SEQUENCE FROM N.A.  
RN Brues M.;  
RN Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
RP [6]  
RA SEQUENCE FROM N.A.  
RN Wieland A., Bonisch H., Bruss M.;  
RT "Molecular cloning of the human and murine high affinity choline  
RT transporters and characterization of the human gene structure.";  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ401466; CAC03171.1; -  
DR EMBL: AF276871; AAC25940.1; -  
DR EMBL: AB043997; BAB18161.1; -  
DR EMBL: AJ308378; CAC88115.1; -  
DR EMBL: AJ308379; CAC88115.1; JOINED.  
DR EMBL: AJ308380; CAC88115.1; JOINED.  
DR EMBL: AJ308381; CAC88115.1; JOINED.  
DR EMBL: AJ308382; CAC88115.1; JOINED.  
DR EMBL: AJ308383; CAC88115.1; JOINED.  
DR EMBL: AJ308384; CAC88115.1; JOINED.  
DR InterPro: IPR001734; Na/solut_symp_3; 1.  
DR Pfam: PF00474; SSF; 1.  
DR PROSITE: PS0283; NA_SOLUT_SYMP_3; 1.  
SQ SEQUENCE 580 AA; 63203 MW; 66CB35496CB6E2D6 CRC64;  
  
Query Match 100.0%; Score 2972; DB 4; Length 580;  
Best Local Similarity 100.0%; Pred. No. 4.3e-205;  
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MAFHVEGLTAAIVFYLLILLVGIWAARWKNSGSAERSEAIIVGGRDILGLVGGFTMTA 60  
Db 1 MAFHVEGLTAAIVFYLLILLVGIWAARWKNSGSAERSEAIIVGGRDILGLVGGFTMTA 60  
  
QY 61 TWGGGYINGTAEAVVPGYGLAWAQAIPGYSLSLILGLFFAKPMRSKGYVTMLDPFQ 120  
Db 61 TWGGGYINGTAEAVVPGYGLAWAQAIPGYSLSLILGLFFAKPMRSKGYVTMLDPFQ 120  
  
QY 121 YGKRMGGLLFPALMGEMFWAAAFSALGATISVIIDVDMHISIIISALATLYLVGG 180  
Db 121 YGKRMGGLLFPALMGEMFWAAAFSALGATISVIIDVDMHISIIISALATLYLVGG 180  
  
QY 181 LYSVAYTDVQLFCIFVGLWISVFPALSHPAVADIGTAVHAKYOKPWLGTVDSEVYSW 240  
Db 181 LYSVAYTDVQLFCIFVGLWISVFPALSHPAVADIGTAVHAKYOKPWLGTVDSEVYSW 240  
  
QY 241 LDSFLLMLGGIPQWQAYFQVRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAISTDW 300  
Db 241 LDSFLLMLGGIPQWQAYFQVRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAISTDW 300  
  
QY 301 NOTAYGLPDKTTEADMILPIVQLCPVYISFFGLGAVSAAYMSSADSSILSASSMFA 360  
Db 301 NOTAYGLPDKTTEADMILPIVQLCPVYISFFGLGAVSAAYMSSADSSILSASSMFA 360  
  
QY 361 RNIIQLSFRQNASDKEIVWVRITVFVGASATAMALLTKTVYGLWYSSDLVYIVFPQ 420  
Db 361 RNIIQLSFRQNASDKEIVWVRITVFVGASATAMALLTKTVYGLWYSSDLVYIVFPQ 420  
  
QY 421 LLCVLFVKGTNTGAVAGYVSGFLRITGGEPLYLQPLIFYPGYPDNGIYNQKPF 480  
Db 421 LLCVLFVKGTNTGAVAGYVSGFLRITGGEPLYLQPLIFYPGYPDNGIYNQKPF 480  
  
QY 481 TLAMVTSFTLNICISYLAKEYLPSGTLPPKLDVDAVARHSEENMDKTLVKNENIKLD 540  
Db 481 TLAMVTSFTLNICISYLAKEYLPSGTLPPKLDVDAVARHSEENMDKTLVKNENIKLD 540  
  
QY 541 ELALVKPROSMTLSSTFTNKEAFLVDSSPEGSGETEDNLQ 580  
Db 541 ELALVKPROSMTLSSTFTNKEAFLVDSSPEGSGETEDNLQ 580  
  
RESULT 2  
Q99PK3  
ID Q99PK3 PRELIMINARY; PRT; 580 AA.  
  
QY 1 MAFHVEGLTAAIVFYLLILLVGIWAARWKNSGSAERSEAIIVGGRDILGLVGGFTMTA 60  
Db 1 MAFHVEGLTAAIVFYLLILLVGIWAARWKNSGSAERSEAIIVGGRDILGLVGGFTMTA 60  
  
QY 61 TWGGGYINGTAEAVVPGYGLAWAQAIPGYSLSLILGLFFAKPMRSKGYVTMLDPFQ 120  
Db 61 TWGGGYINGTAEAVVPGYGLAWAQAIPGYSLSLILGLFFAKPMRSKGYVTMLDPFQ 120  
  
QY 121 YGKRMGGLLFPALMGEMFWAAAFSALGATISVIIDVDMHISIIISALATLYLVGG 180  
Db 121 YGKRMGGLLFPALMGEMFWAAAFSALGATISVIIDVDMHISIIISALATLYLVGG 180  
  
QY 181 LYSVAYTDVQLFCIFVGLWISVFPALSHPAVADIGTAVHAKYOKPWLGTVDSEVYSW 240  
Db 181 LYSVAYTDVQLFCIFVGLWISVFPALSHPAVADIGTAVHAKYOKPWLGTVDSEVYSW 240  
  
QY 241 LDSFLLMLGGIPQWQAYFQVRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAISTDW 300  
Db 241 LDSFLLMLGGIPQWQAYFQVRLSSSSATYAQVLSFLAAGCLVMAIPAILIGAISTDW 300  
  
QY 301 NOTAYGLPDKTTEADMILPIVQLCPVYISFFGLGAVSAAYMSSADSSILSASSMFA 360  
Db 301 NOTAYGLPDKTTEADMILPIVQLCPVYISFFGLGAVSAAYMSSADSSILSASSMFA 360  
  
QY 361 RNIIQLSFRQNASDKEIVWVRITVFVGASATAMALLTKTVYGLWYSSDLVYIVFPQ 420  
Db 361 RNIIQLSFRQNASDKEIVWVRITVFVGASATAMALLTKTVYGLWYSSDLVYIVFPQ 420  
  
QY 421 LLCVLFVKGTNTGAVAGYVSGFLRITGGEPLYLQPLIFYPGYPDNGIYNQKPF 480  
Db 421 LLCVLFVKGTNTGAVAGYVSGFLRITGGEPLYLQPLIFYPGYPDNGIYNQKPF 480  
  
QY 481 TLAMVTSFTLNICISYLAKEYLPSGTLPPKLDVDAVARHSEENMDKTLVKNENIKLD 540  
Db 481 TLAMVTSFTLNICISYLAKEYLPSGTLPPKLDVDAVARHSEENMDKTLVKNENIKLD 540  
  
QY 541 ELALVKPROSMTLSSTFTNKEAFLVDSSPEGSGETEDNLQ 580  
Db 541 ELALVKPROSMTLSSTFTNKEAFLVDSSPEGSGETEDNLQ 580  
  
RESULT 3  
Q99PK3  
ID Q99PK3 PRELIMINARY; PRT; 580 AA.
```

DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE	High affinity choline transporter.
GN	SLC5A7 OR CHT1.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RN	SEQUENCE FROM N.A.
RP	STRAIN=BALB/CJ; TISSUE=BRAIN STEM;
RC	Bruss M.;
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN	[2]
RN	SEQUENCE FROM N.A.
RP	STRAIN=BALB/CJ; TISSUE=BRAIN STEM;
RC	Wieland A., Bonisch H., Bruss M.;
RT	"Molecular cloning of the human and murine high affinity choline
RT	transporters and characterization of the human gene-structure.";
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL: AJ401467; CAC03719.1; -
DR	MGI: 1927126; Slc5a7.
DR	MGD: MGI:1927126; Na/solut_sympot.
DR	InterPro: IPR001734; Na/solut_sympot.
DR	Pfam: PF00474; SSF: 1.
DR	PROSITE: PS0283; NA_SOLUT_SYMP_3; 1.
DR	SEQUENCE 580 AA; 63331 MW; A4F1387CA9AAFE CRC64;
QY	Query Match 93.9%; Score 2791; DB 11; Length 580;
QY	Best Local Similarity 92.4%; Pred. No. 4.2e-192;
QY	Matches 536; Conservative 24; Mismatches 20; Indels 0; Gaps
QY	1 MAFHVEGLIAIVFYLLLLVGIWAARPTKNSGAERSEAIIVGGRDGLLVGGFTMTA 60
QY	1 MSFHVEGLVAIFLLIFLVGIWAARKTKNSGNPEHSEAIIVGGRDGLLVGGFTMTA 60
QY	61 TWVGGYINGTAAVVPVPGYGLAWAQAQPGYSLSLILGGLFFAKPMRSGYVMTLDPFQ 120
QY	61 TWVGGYINGTAAVVPVPGYGLAWAQAQPGYSLSLILGGLFFAKPMRSGYVMTLDPFQ 120
QY	121 IYKRMGGLLFTPALMGEMFMAAIFISALGATISVIIDVDMHLSVIAIALIATLYLVGG 180
QY	121 IYKRMGGLLFTPALMGEMFMAAIFISALGATISVIIDVDMHLSVIAIALIATLYLVGG 180
QY	181 LYSVATDVVQLFCIFVGLWISVPFALSHPAVDIGFTAVHAKYQKPLGTVDSDSEYVS 240
QY	181 LYSVATDVVQLFCIFVGLWISVPFALSHPAVDIGFTAVHAKYQKPLGTVDSDSEYVS 240
QY	241 LDSFLLMLGGIPQWAFQRYLVSSSSATYAQVLSFLAAGCLVMAIPAILIGATGASTD 300
QY	241 LDNFLLMLGGIPQWAFQRYLVSSSSATYAQVLSFLAAGCLVMAIPAILIGATGASTD 300
QY	301 NQATYGLPDKPTTEADMILPIVQLYLCPPVVISFGLGAVSAAYMSADSSILSASSMFA 360
QY	301 NQATYGLPDKPTTEADMILPIVQLYLCPPVVISFGLGAVSAAYMSADSSILSASSMFA 360
QY	361 RNIIQLSFRONASDKIEIWMNRITVVFEGASATAMALLTKTVGLWYSSDLVYIVFPQ 420
QY	361 RNIIQLSFRONASDKIEIWMNRITVVFEGASATAMALLTKTVGLWYSSDLVYIVFPQ 420
QY	421 LLCVLFVKGNTYCAVAGYVSGFLRTIRGGPPYLYLOPLIFPGYPDPDNGIYNQKFFPK 480
QY	421 LLCVLFVKGNTYCAVAGYVSGFLRTIRGGPPYLYLOPLIFPGYPDPDNGIYNQKFFPK 480
QY	481 TLAMVTSPLTNCISYLAKYLFESEGLTPPKLDVDFAVVARHSEENKDKTILVKNENIKLD 540
QY	481 TLAMVTSPLTNCISYLAKYLFESEGLTPPKLDVDFAVVARHSEENKDKTILVKNENIKLD 540
QY	541 ELALVKPQSMSTLSTFTTNKEAFLDQVDSSEGSGETDNLQ 580
QY	541 ELAPVKPQSMSTLSTFTTNKEAFLDQVDSSEGSGETDNLQ 580
QY	RESULT 5

Q99PK3;	01-JUN-2001 (TReMBLrel. 17, Created)
Q99PK3;	01-JUN-2001 (TReMBLrel. 17, Last sequence update)
Q99PK3;	01-JUN-2002 (TReMBLrel. 21, Last annotation update)
Q99PK3;	Sodium and chloride-dependent high-affinity choline transporter.
Q99PK3;	SLC5A7.
Q99PK3;	NCBI_TaxID=10090;
Q99PK3;	[1]
Q99PK3;	SEQUENCE FROM N.A.
Q99PK3;	Apparsundaram S.A., Ferguson S.M., Blakely R.D.;
Q99PK3;	"Molecular cloning and characterization of human and murine high-
Q99PK3;	affinity choline transporters";
Q99PK3;	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
Q99PK3;	EMBL; AF276872; AAG36945.2; -
Q99PK3;	MGD; MGI:1927126; SLC5A7.
Q99PK3;	InterPro; IPR001734; Na/solut_symport.
Q99PK3;	Pfam; PF00474; SSP; 1.
Q99PK3;	PROSITE; PS50283; NA_SOLUT_SYMPT 3; 1.
Q99PK3;	SEQUENCE 580 AA; 63383 MW; DDBF58ED428270AF CRC64;
Q99PK3;	Query Match 94.0%; Score 2795; DB 11; Length 580;
Q99PK3;	Best Local Similarity 92.6%; Pred. No. 2.1e-192;
Q99PK3;	Matches 537; Conservative 23; Mismatches 20; Indels 0; Gaps 0;
Q99PK3;	1 MAPHVEGLIAIVFYLLILLVGIWAARWTKNSGASERSEAIIVGGRDGLLVGGFTMTA 60
Q99PK3;	1 MPFHVEGLVAILFYLLIFLVGIWAARWTKNSGNPEERSEAIIVGGRDGLLVGGFTMTA 60
Q99PK3;	61 TWVGGGYINGTAEAVYVPGYGLAWAQPIGYSLSLILGGLFFAKPMRSGYVMTLDPFQ 120
Q99PK3;	61 TWVGGGYINGTAEAVYVPGYGLAWAQPIGYSLSLILGGLFFAKPMRSGYVMTLDPFQ 120
Q99PK3;	121 IYGRMGGLLIPALMGEMFWAAIFPALTGATISVIIDVDHISVTISALIAITYLVGG 180
Q99PK3;	121 IYGRMGGLLIPALMGEMFWAAIFPALTGATISVIIDVDHISVTISALIAITYLVGG 180
Q99PK3;	181 LYSVAYTDVQLFCIFVGLMTISVPFALSHPAVADIGFTAVHAKYQKPLGTVDSSEVSW 240
Q99PK3;	181 LYSVAYTDVQLFCIFVGLMTISVPFALSHPAVADIGFTAVHAKYQKPLGTVDSSEVSW 240
Q99PK3;	241 LDSFLLMLGGIPWQAYFORVLSSSSATTYAQVLSFLAAGCLVNAIPAILGAGSTDW 300
Q99PK3;	241 LDFLLMLGGIPWQAYFORVLSSSSATTYAQVLSFLAAGCLVNAIPAILGAGSTDW 300
Q99PK3;	301 NOTAYGLDPDKTTEADMLPIVLOYLCPVVISFEGLCANSAVMSADSSILSASSMFA 360
Q99PK3;	301 NOTAYGLDPDKTTEADMLPIVLOYLCPVVISFEGLCANSAVMSADSSILSASSMFA 360
Q99PK3;	361 RNIYQLSPQNADSKIEIVVMRITVVFEGASATAMALLTKTVYGLWLSDDLVIIVFPQ 420
Q99PK3;	361 RNIYQLSPQNADSKIEIVVMRITVVFEGASATAMALLTKTVYGLWLSDDLVIIVFPQ 420
Q99PK3;	421 LLCVLFVKGTNTYGVAGVYVSGLFIRITGGEPYLYQLPIFYPGYIPDDNGIYNQKPFK 480
Q99PK3;	421 LLCVLFVKGTNTYGVAGVYVSGLFIRITGGEPYLYQLPIFYPGYIPDDNGIYNQKPFK 480
Q99PK3;	481 TLAMVTSFTNTICISYLAKYLFESGTLPPKLDVDFAVVARHSEENKMTILVKENIKLD 540
Q99PK3;	481 TLSMVTSTFTNTICISYLAKYLFESGTLPPKLDVDFAVVARHSEENKMTILVKENIKLD 540
Q99PK3;	541 ELALVPRQSLTSSFTTNKEAFLDVDDSSPEGSGTEDNLQ 580
Q99PK3;	541 ELALVPRQSLTSSFTTNKEAFLDVDDSSPEGSGTEDNLQ 580
Q99PK3;	RESULT 4
Q99PK3;	ID Q9ESW5 PRELIMINARY; PRT; 580 AA.
Q99PK3;	AC Q9ESW5
Q99PK3;	01-MAR-2001 (TReMBLrel. 16, Created)

C48D1.3 protein (High-affinity choline transporter CHO-1).

C48D1.3 OR CHO-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
ON NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Burton J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20116099; PubMed=9851916;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RX MEDLINE=20116099; PubMed=10649566;
RA Okuda T., Haga T., Kanai Y., Endou H., Ishihara T., Katsura I.;
RT "Identification and characterization of the high-affinity choline
transporter.";
RL Nat. Neurosci. 3:120-125(2000).
DR EMBL; Z81049; CAB02847.2; .
DR EMBL; AB030946; BAA90483.1; .
DR InterPro; IPR001734; Na/solut_sympot.
DR Pfam; PF00474; SSF; 1.
DR PROSITE; PS0283; NA_SOLUT_SYMP_3; 1.
SQ SEQUENCE 576 AA; 62427 MW; FAB09778358288D9 CRC64;

DE

GN

OS

OC

OC

OC

ON

RN

RP

RA

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

01-MAR-2001 (TREMBlrel. 16, Created)

01-MAR-2001 (TREMBlrel. 16, Last sequence update)

01-MAR-2002 (TREMBlrel. 20, Last annotation update)

Choline cotransporter.

OS Limulus polyphemus (Atlantic horseshoe crab).

OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;

OC Limulidae; Limulus.

ON NCBI_TaxID=6850;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21261948; PubMed=11368908;

RA Wang Y., Cao Z., Newkirk R.F., Ivy M.T., Townsel J.G.;

RT "Molecular cloning of a cDNA for a putative choline co-transporter

from Limulus CNS.";

RL Gene 268:123-131(2001).

DR EMBL; AY011119; AAG41055.1; .

DR InterPro; IPR001734; Na/solut_sympot.

DR Pfam; PF00474; SSF; 1.

DR PROSITE; PS0283; NA_SOLUT_SYMP_3; 1.

DR SEQUENCE 579 AA; 62937 MW; FE7F29D4FAF47F04 CRC64;

Query Match 51.5%; Score 1530.5; DB 5; Length 579;

Best Local Similarity 52.0%; Pred. No. 1.1e-101;

Matches 305; Conservative 115; Mismatches 134; Indels 33; Gaps 11;

QY 1 MAFHVEGLIAIVFYLLILVGINAAWRTKNGSAAERSEALIVGGRDIGLVGGFTMTA 60

DB 1 MAVNILGVVSGIFVILVGINAS-RKKTSQSGSETEIMLAGRNIGLVGLTMTA 59

QY 61 TWVGGYINGTAEAVVPGYGLAWAQAIPGYSLILGLLFAKPMRSKGYTMDLPQQ 120

DB 60 TWVGGYINGTAEAVV--NNGLVWCAQAPFGYALSIFGIGVFAKMRSGQYVMTDLPQ 117

QY 121 IVGRMGGLLFTPALMGEMFWAAAFSALGATISVIIDVDHISVIALIATLTVLGG 180

DB 118 NFGRMGGLLFTPALMGEMFWAAAFSALGATISVIIDVDHISVIALIATLTVLGG 177

QY 181 LYSVATDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPLGTVDSDSEVSW 240

DB 178 FYSVATDVVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPLGTVDSDSEVSW 229

QY 241 LQSFLLMLGGIPWQAYFORVLSSSATVAQVLSFLAAGCLVMAIPAILIGAIGASTDW 300

DB 230 VDTWLLIFGGIPWQAYFORVLSAKNSVNAQVLSYVAAGCVVMAIPAILIGVIAKATAW 289

QY 301 NOTAYGLPDPKTEEAADMILPIVQLVCPVYISFFGLGAVSAVMSADSSILSASSMFA 360

DB 290 NETALGN--PLTPNDTSLVPLVHLVLTPTAVSFFGLGAVSAVMSADSSILSASSLFS 347

QY 361 RNIIQLSFRONASDKKEIVWVRITVVFEGASATAMALLTIVYGLWYLSDDLVIIVFPQ 420

DB 348 RNVYKLFQKASEREVWVIRISILVVGILATAMALTIVKSVGLWYLSDDLVIIVFPQ 407

QY 421 LLCVLFVKG--TNTYGVAGVSGFLRLITGGEPYLYLOPLIFPGYPPDNGIYNQKPPF 479

DB 408 LLCVHLKKYNTYVGLSAYIVGFLRALGESITGLPEVIHYF--FFSETSG---QRPFF 463

QY 480 KTLAWVTSFTNLCISYLAKEYLFPESGTLPLPKLDFVAVARHSEENMDKTLVKNENIK- 538

DB 464 RTLNLASLTLLAISGITKWFEMNHLPAKLDIFRCVT--NIDEN-----IILKIQKLOG 516

QY 539 -----LDEL--ALVKPRQSMSTSTFTNKEAFLVDSDSPESGSGEDN 578

DB 517 GAMPVLDISIKKEIYQKDNNSFNTVNSGNAELTDTSTYSGKIKNN 563

RESULT 8

ID 002228 PRELIMINARY; PRT; 576 AA.

AC 002228; Q0NL58;

DT 01-JUL-1997 (TREMBlrel. 04, Created)

DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE

GN

OS

OC

OC

ON

RN

RP

RA

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

RA

RT

RL

RN

RP

RC

RX

```

Query Match      24.3%   Score 723.5; DB 5; Length 278;
Best Local Similarity 64.2%; Pred. No. 3.5e-44;
Matches 136; Conservative 33; Mismatches 38; Indels 5; Gaps 3;
QY 4 HVEGLAIIVLLILLVGIWAAWTKNSGAERSEALIIVGGRDIGLVGFGFWMTATW 63
    :: ::::|||||:||||| | || | ::: | |||| | |||||

```


Db	410	Y-TKIRKASEKELMIAGRFLFVLVLIGISIAWPIVQSAQSQGLFDYIQSITSYLGPPIA	468
QY	419	PQLLCVLFVKQNTYCAVAGYVSGFLRI-----TG-----GEPVLY	455
Db	469	AVFLAAIFCKRVNEQGAFLGLILGLFELGISRMITEFAYGTGSCMPSPNCPTKICGVHLY	528
QY	456	LOPLFIYPGYDDNGIYNQKFPFKTAMVTSFLTNCISYLAK	499
Db	529	FAILF-----VISVITILILSFLTK	549
RESULT	15		
ID	Q8WXX8	PRELIMINARY; PRT; 675 AA.	
AC	Q8WXX8;		
DC	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Sodium/glucose cotransporter KSTL.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Roll P., Massacrier A., Pereira S., Robaglia-Schlupp A., Cau P.,		
RA	Szepietowski P.;		
RT	"New human sodium glucose cotransporter gene (hKSTL).";		
RL	Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.		
DR	EMBL: AY044906; AAK97784.1; -		
DR	InterPro: IPR001734; Na/solut_sympot.		
DR	Pfam: PF00474; SSF: 1.		
DR	TIGRFAMS; TIGR00813; sss: 1.		
DR	PROSITE; P50283; NA_SOLUT_SYMP_3; 1.		
SO	SEQUENCE 675 AA; 74036 MW; 44F7BA6D2FE92335 CRG64;		

Qy	408	LSSDLVI-----VIPOLLCLVLFVKCTNTYGAAGVGSVSLFLRLTG-GEPLYLQPLIF	461
Db	451	IQSISSYLQPPVAVF---TMCGFWRKTNKEGAFGLISGLLGLVRLVLDIFYVQPR-	506
Qy	462	YGYYPDNDGIYNQKPEFKTLMWTSPLTNICISYLAKLYFESGTLPKKLDV-----	513
Db	507	---DQDERPVLKSHYLFSEMLSTVTLITVTSWSF-----TEPPSKEMVSHTWFT	558
Qy	514	-FDAAVVAHSEENMDKTILVKNENIKLD-----ELALVKPQSGMTLSLSTFNKEA	562
Db	559	RHPVVKQKGAAPPAALSLTSLONGMPEASSSSSVQFMVQENTSKTHSCDMTPKGS	615

Search completed: March 31, 2003, 18:37:43
Job time : 69 secs

us-10-069-541-6.rsp

Tue Apr 1 13:57:27 2003

p94408 bacillus su
O67658 aquifex aeo
P15584 parametium
P29926 paracoccus
P30143 escherichia
P48920 chondrus cr
P98004 sulfolobus
P06264 sulfolobus
P06264 sulfolobus
P75783 escherichia
P54571 bacillus su
Q08579 saccharomyc

34 132 4.4 492 1 YCLF_BACSU
35 132 4.4 499 1 MYIN_AQUAE
36 131 4.4 570 1 NU5M_PARDE
37 125.5 4.2 499 1 NOOE_PARDE
38 125 4.2 476 1 YAAJ_ECOLI
39 124.5 4.2 666 1 NU5M_CHOCR
40 123 4.1 517 1 OOKL_SULAC
41 120 4.1 692 1 NUSC_MARPO
42 120 4.0 687 1 CSTA_HELPJ
43 120 4.0 741 1 YBIO_ECOLI
44 119 4.0 468 1 YOKI_BACSU
45 119 4.0 599 1 THYI_YEAST

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

March 31, 2003, 17:49:41 ; Search time 38 Seconds

(without alignments)
633.060 Million cell updates/sec

Title: US-10-069-541-6

Perfect score: 2972

Sequence: 1 MAFHVEGLIAIIVYLLILL.....EAFLDVDSPEGSGETDNLQ 580

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	308.5	10.4	662	1 SL51_RABIT	P11170 oryctolagus
2	308	10.4	670	1 SL52_RAT	P53792 rattus norv
3	306	10.3	664	1 SL51_HUMAN	P13866 homo sapien
4	306	10.3	665	1 SL51_RAT	P53790 rattus norv
5	303.5	10.2	660	1 SL54_PIG	P31636 sus scrofa
6	298	10.0	672	1 SL52_RABIT	P26430 oryctolagus
7	294	9.9	659	1 SL54_HUMAN	Q9ny91 homo sapien
8	294	9.9	664	1 SL51_SHEEP	P53791 ovis aries
9	293.5	9.9	543	1 SGLT_VIBPA	P96169 vibrio para
10	292	9.8	672	1 SL52_HUMAN	P31639 homo sapien
11	290	9.8	656	1 SL54_MOUSE	Q9et37 mus musculu
12	289	9.7	718	1 SL53_MOUSE	Q91k22 mus musculu
13	285	9.6	492	1 OPUF_BACSU	O06493 bacillus su
14	278.5	9.4	718	1 SL53_BOVIN	P31637 canis fami
15	275	9.3	718	1 SL53_HUMAN	P53793 bos taurus
16	272.5	9.2	718	1 SL53_HUMAN	P53794 homo sapien
17	260.5	8.8	504	1 PUTP_HAEIN	P45174 haemophilus
18	260.5	8.8	504	1 SL51_PIG	P26429 sus scrofa
19	253	8.5	502	1 PUTP_ECOLI	P07117 escherichia
20	249	8.4	502	1 PUTP_SALTY	P10502 salmonella
21	239.5	8.1	484	1 PANF_HAEIN	P44963 haemophilus
22	239	8.0	643	1 YJCG_ECOLI	P32705 escherichia
23	231	7.8	643	1 SL55_HUMAN	Q92911 homo sapien
24	229.5	7.7	636	1 SL56_RABIT	Q9xt77 oryctolagus
25	226	7.6	571	1 YIDK_ECOLI	P31448 escherichia
26	219.5	7.4	634	1 SL56_RAT	O70247 rattus norv
27	218	7.3	618	1 SL55_RAT	Q63008 rattus norv
28	209	7.0	635	1 SL56_HUMAN	Q9y289 homo sapien
29	206.5	6.9	735	1 PANF_ECOLI	P16256 escherichia
30	198.5	6.7	735	1 DUR3_YEAST	P33413 saccharomyc
31	198	6.7	516	1 YWCA_BACSU	P39599 bacillus su
32	137	4.6	657	1 NU5M_EMENI	P11628 emericella
33	132	4.4	482	1 ARCD_PSEAE	P18275 pseudomonas

ALIGNMENTS

RESULT 1

SL51_RABIT STANDARD; PRT; 662 AA.
ID SL51_RABIT
AC P11170; 1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JUL-1989 (Rel. 36, Last annotation update)
DE Sodium/glucose cotransporter 1 (Na⁺/glucose cotransporter 1)
DE (High affinity sodium-glucose cotransporter).
DE SLC5A1 OR SGLT1.
GN Oryctolagus cuniculus (Rabbit).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white;
RX MEDLINE=88065856; PubMed=2446136;
RA Hediger M.A., Coady M.J., Ikeda T.S., Wright E.M.;
RT "Expression cloning and cDNA sequencing of the Na⁺/glucose co-transporter."
RL Nature 330:379-381(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Kidney cortex;
RX MEDLINE=91223090; PubMed=2025641;
RA Morrison A.I., Panayotova-Heller M., Feigl G., Schoelermann B.,
RA Kinne R.K.H.;
RT "Sequence comparison of the sodium-D-glucose cotransport systems in rabbit renal and intestinal epithelia."
RL Biochim. Biophys. Acta 1089:121-123(1991).
CC -!- FUNCTION: ACTIVELY TRANSPORTS GLUCOSE INTO CELLS BY NA⁺ CO-TRANSPORT WITH A NA⁺ TO GLUCOSE COUPLING RATIO OF 2:1.
CC -!- FUNCTION: EFFICIENT SUBSTRATE TRANSPORT IN MAMMALIAN KIDNEY IS PROVIDED BY THE CONCENTRATED ACTION OF A LOW AFFINITY HIGH CAPACITY AND A HIGH AFFINITY LOW CAPACITY NA⁺/GLUCOSE COTRANSPORTER.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: FOUND PREDOMINANTLY IN INTESTINE AND IN OUTER RENAL MEDULLA.
CC -!- DISEASE: MUTATION OF ASP-28 IS IMPLICATED IN GLUCOSE/GALACTOSE MALABSORPTION.
CC -!- SIMILARITY: BELONGS TO THE SODIUM: Solute symporter family (SSF).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X06419; CAA29727.1; -
CC EMBL; X55355; CAA39040.1; -
CC PIR; S00515; S00515.
CC PIR; S15974; S15974.
CC -----

QY	437	AGYVSGFLRI	-----TG-----	-----CEPVLVLOPLFVPGYVDDNGIY	473
Db	487	WGLVGLGIGIRSMIT	FAVCTGSCMEPNSCPTICGVHYLYFAILF	-----	534
QY	474	NQKFFFTLAWVSFL	TNLCISYIAKYLFSGTLPKPLVDFAVA	-RHSEENMDKTLV	532
Db	535	-----	-----VISIITVVVSFLTKEPI	-----PDVHLYRLCWSLRNSKE	-----568
QY	533	KNENIKLD	-ELALVKPRQSMTSSFTTKKEAF	-----LDVDSPEGSGTGD	577
Db	569	--ERIDIDAGEEDIQ	EAPAEATDEYVKKKGFFRAYDLFCGLDODKGP	KMTKEE	623
RESULT 2					
SL52_RAT	ID	SL52_RAT	STANDARD;	PRT;	670 AA.
AC	P53792;				
DT	01-OCT-1996	(Rel. 34, Created)			
DT	01-OCT-1996	(Rel. 34, Last sequence update)			
DT	15-JUL-1998	(Rel. 36, Last annotation update)			
DE	Sodium/glucose cotransporter 2 (Na ⁺ /glucose cotransporter 2)				
DE	(Low affinity sodium-glucose cotransporter).				
GN	SLC5A2 OR SGT2.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPRAIN-Sprague-Dawley; TISSUE=Kidney;				
RC	MEDLINE=36094332; PubMed=7493971;				
RA	You G., Lee W.-S., Barros E.J.G., Kanai Y., Huo T.-L., Khawaja S.,				
RA	Wells R.G., Nigam S.K., Hediger M.A.;				
RT	"Molecular characteristics of Na ⁺ -coupled glucose transporters in				
RT	adult and embryonic rat kidney".				
RL	J. Biol. Chem. 270:29365-29371(1995).				
CC	-!- FUNCTION: SODIUM-DEPENDENT GLUCOSE TRANSPORTER. HAS A NA+ TO				
CC	GLUCOSE COUPLING RATIO OF 1:1.				
CC	-!- FUNCTION: EFFICIENT SUBSTRATE TRANSPORT IN MAMMALIAN KIDNEY IS				
CC	PROVIDED BY THE CONCERTED ACTION OF A LOW AFFINITY HIGH CAPACITY				
CC	AND A HIGH AFFINITY LOW CAPACITY NA ⁺ /GLUCOSE COTRANSPORTER				
CC	ARRANGED IN SERIES ALONG KIDNEY PROXIMAL TUBULES.				
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein.				
CC	-!- TISSUE SPECIFICITY: KIDNEY, IN PROXIMAL TUBULE S1 SEGMENTS.				
CC	DEVELOPMENTAL STAGE: APPEARS ON EMBRYONIC DAY 17 AND GRADUALLY				
CC	INCREASES UNTIL DAY 19. DECREASES BETWEEN DAY 19 AND GRADUALLY				
CC	-!- PTM: GLYCOSYLATED AT A SINGLE SITE.				
CC	-!- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).				
CC	-----				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; U29881; AAC52325.1; -				
DR	InterPro: IPR001734; Na/solut_symport.				
DR	Pfam: PF00474; SSF; 1.				
DR	TIGRFAMS; TIGR00813; sss; 1.				
DR	PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.				
DR	PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.				
DR	PROSITE; PS50283; NA_SOLUT_SYMP_3; 1.				
KW	Transport; Sugar transport; Transmembrane; Sodium transport; Symport;				
FT	glycoprotein.				
FT	DOMAIN	1	23	CYTOPLASMIC (POTENTIAL).	
FT	TRANSMEM	24	42	POTENTIAL.	
FT	DOMAIN	43	59	EXTRACELLULAR (POTENTIAL).	
FT	TRANSMEM	60	80	POTENTIAL.	
FT	DOMAIN	81	100	CYTOPLASMIC (POTENTIAL).	
FT	TRANSMEM	101	121	POTENTIAL.	

[5]
 VARIANT GGM GLY-28.
 PubMed-8195156;
 Turk E., Martin M.G., Wright E.M.;
 "Structure of the human Na⁺/glucose cotransporter gene SGLT1.";
 J. Biol. Chem. 269:15204-15209(1994).
 - FUNCTION: ACTIVELY TRANSPORTS GLUCOSE INTO CELLS BY NA⁺
 CO-TRANSPORT WITH A NA⁺ TO GLUCOSE COUPLING RATIO OF 2:1.
 - PROVIDED BY THE CONCENTRATED ACTION OF A LOW AFFINITY HIGH CAPACITY
 AND A HIGH AFFINITY LOW CAPACITY NA(+)/GLUCOSE COTRANSPORTER
 ARRANGED IN SERIES ALONG KIDNEY PROXIMAL TUBULES.
 - SUBCELLULAR LOCATION: Integred membrane protein.
 - TISSUE SPECIFICITY: Expressed mainly in intestine and kidney.
 - DISEASE: Defects in SLC5A1 are the cause of the intestinal
 monosaccharide transporter deficiency known as congenital glucose-
 galactose malabsorption (GGM). It is an autosomal recessive
 disorder manifesting itself within the first weeks of life. It is
 fatal unless glucose and galactose are eliminated from the diet.
 - SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; L29339; AAB59448.1; JOINED.
 EMBL; L29328; AAB59448.1; JOINED.
 EMBL; L29330; AAB59448.1; JOINED.
 EMBL; L29329; AAB59448.1; JOINED.
 EMBL; L29331; AAB59448.1; JOINED.
 EMBL; L29332; AAB59448.1; JOINED.
 EMBL; L29333; AAB59448.1; JOINED.
 EMBL; L29334; AAB59448.1; JOINED.
 EMBL; L29335; AAB59448.1; JOINED.
 EMBL; L29336; AAB59448.1; JOINED.
 EMBL; L29337; AAB59448.1; JOINED.
 EMBL; L29338; AAB59448.1; JOINED.
 EMBL; M24847; AAB59448.1; JOINED.
 EMBL; AL022321; CAAB1845.2; -
 EMBL; Z83849; CAB06087.1; JOINED.
 EMBL; Z74021; CAAB06087.1; JOINED.
 EMBL; Z80998; CAB02632.2; -
 EMBL; Z83839; CAB06087.1; JOINED.
 EMBL; Z74021; CAB06087.1; JOINED.
 EMBL; Z80998; CAB06087.1; JOINED.
 EMBL; Z83849; CAB06087.1; JOINED.
 EMBL; AL022321; CAB06087.1; JOINED.
 PIR; A33545; A33545.
 Genew; HGNC:11036; SLC5A1.
 MIM; 182380; -
 InterPro; IPR001734; Na/solut_symport.
 Pfam; PF00474; SSF; 1
 TIGRFAMs; TIGR00813; Sss; 1.
 PROSITE; PS00456; NA_SOLUT_SYM_1; 1.
 PROSITE; PS00457; NA_SOLUT_SYM_2; 1.
 PROSITE; PS0283; NA_SOLUT_SYM_3; 1.
 Transport; Sugar transport; Transmembrane; Sodium transport; Symport;
 Glycoprotein; Disease mutation.
 FT DOMAIN 1 28 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 29 47 POTENTIAL.
 FT DOMAIN 48 64 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 65 85 POTENTIAL.
 FT DOMAIN 86 105 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 106 126 POTENTIAL.
 FT DOMAIN 127 171 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 172 191 POTENTIAL.
 FT DOMAIN 193 208 CYTOPLASMIC (POTENTIAL).
 RESULT 4
 SL51_RAT
 ID SL51_RAT
 AC P53790; P97787; STANDARD; PRT; 665 AA.

FT TRANSMEM 209 229 POTENTIAL.
 FT DOMAIN 230 270 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 271 291 POTENTIAL.
 FT DOMAIN 292 314 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 315 334 POTENTIAL.
 FT DOMAIN 335 423 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 424 443 POTENTIAL.
 FT DOMAIN 444 455 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 456 476 POTENTIAL.
 FT DOMAIN 477 526 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 527 547 POTENTIAL.
 FT DOMAIN 548 642 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 643 663 POTENTIAL.
 FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
 FT VARIANT 28 28 D -> G (IN GGM).
 FT VARIANT 28 28 /FTID-VAR_013630.
 FT VARIANT 28 28 D -> N (IN GGM).
 FT VARIANT 28 28 /FTID-VAR_007168.
 SQ SEQUENCE 664 AA; 73497 MW; 2B403376595EAB74 CRC64;
 Query Match 10.3%; Score 306; DB 1; Length 664;
 Best Local Similarity 22.8%; Pred. No. 7.4e-13;
 Matches 148; Conservative 104; Mismatches 218; Indels 178; Gaps 30;
 QY 11 IIVFYLLILLVGINAWRTKNSGAERSEAIIVGGRDGLLVGGFTMTATWVGQYING 70
 DB 32 IIVFYVYVMAVGLWAMFST-NRGTV- -GGFFLAGRSVMVWPVIGASLFASNGHFVYG 86
 QY 71 TAEAVYVPGYGLAWAQAIPGYS- -LSLILGGFFAKPMRSK-GYVTMLDPPQIYQK 124
 DB 87 LA- -GTCAASGAIAGGFENALVVLVGLWLFV-PIYKAGVVTM- -PEYLRLK 134
 QY 125 RMGG- -LLFIPALMGEMWAAAFISALCATISVIIDDMHLSVIALIA 172
 DB 135 RFGQRIQVYLSLSLLLYITFTKISADIFSCAIF- -INLALGMLYLAIFLLAIT 188
 QY 173 TLYTLVGLXSVAYTDVQVLCIFVGLWISVPPFALSHPAVADIGTAVHAKYK- -PWL- 229
 DB 189 ALYITITGGLAAVITDPLQIVMLVGLSLITGFAFHEVG- -GYDAFMKYMKAIPITV 244
 QY 230 - -GTVDSESVYS-WLDSFLL- -MLGGIPW- - - - -QAYFORVLS 264
 DB 245 SDGNTTFQEKYTPRADSFHIFRDLTGLDLPWPGFIFGMSILTLWYCTQIVIVQRCLSA 304
 QY 265 SSATYAO- -VLSFLAAGCLVMAIPAIL- - - - -GASTDWNQT 303
 DB 305 KMSHVKGCCILCGYLKLMPTMVPVMSIRILYTEKACVVPSECEKCYGKVGCTNI 364
 QY 304 AYGLDPDKTTEADMTLPVQLCPVYISFFGLGAVSAVMSADSSILSASSMFARNI 363
 DB 365 AY- - - - -PTLVVLPMPNGRLGLMSVMLASLMSLSLTSIFNSASTLFTMDI 409
 QY 364 YQLSPQNASKDEIYVVMRITVVFV-FGASATAMALLTKTVYG- -LWYLSDDLVI- -VIF 418
 DB 410 Y-AKVRKRASEKELMIAGRLFTLVILIGISIAWVIVQSQQLFDYIQSITSYLGPIIA 468
 QY 419 POLLCLVFKGNTYGVAGVYVSGFLRI- - - - -TG- - - - -GEPVLY 455
 DB 469 AVFLAIFWKRVPNEPGAFWGLIILIGISRMITEFAFGTSCMEPSNCPITICGVHLY 528
 QY 456 LQPIIFYGYPDDNGIYNQKFPKTLAMVTSFLTNTICISVLAKYLFESGLTPKLDVDF 515
 DB 529 FAILF- - - - -AISFTIIVISLLTKPI- - - - -PDVHLR 558
 QY 516 AV- -VARHSEENMDKTLVKNENIKLDELALVKPQSMTSLSTFNKE 561
 DB 559 LCWSLANSKEERID- -LDAEENIQ- - - - -EGPKETIETQVPEKK 598
 RESULT 4
 SL51_RAT
 ID SL51_RAT
 AC P53790; P97787; STANDARD; PRT; 665 AA.

Tue Apr 1 13:57:27 2003

01-OCT-1996 (Rel. 34, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 Sodium/glucose cotransporter 1 (Na(+)/glucose cotransporter 1)
 (High affinity sodium-glucose cotransporter).
 SLC5A1 OR SGLT1.
 Rattus norvegicus (Rat).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=Sprague-Dawley; TISSUE=Kidney;
 MEDLINE=94216314; PubMed=8163506;
 Lee W.S., Kanai Y., Wells R.G., Hediger M.A.;
 "The high affinity Na(+)/glucose cotransporter. Re-evaluation of
 function and distribution of expression.";
 J. Biol. Chem. 269:12032-12039(1994).
 [2]
 SEQUENCE FROM N.A.
 Kasahara M., Mori K.;
 Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
 [3]
 SEQUENCE FROM N.A.
 STRAIN=Sprague-Dawley; TISSUE=Jejunum;
 Aoshima H., Yokoyama T., Tanizaki J., Izu H., Yamada M.;
 "The sugar specificity of Na/glucose cotransporter from rat jejunum.";
 Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 -1- FUNCTION: ACTIVELY TRANSPORTS GLUCOSE INTO CELLS BY NA+
 CO-TRANSPORT WITH A NA+ TO GLUCOSE COUPLING RATIO OF 2:1.
 -1- FUNCTION: EFFICIENT SUBSTRATE TRANSPORT IN MAMMALIAN KIDNEY IS
 PROVIDED BY THE CONCENTRATED ACTION OF A LOW AFFINITY HIGH CAPACITY
 AND A HIGH AFFINITY LOW CAPACITY NA(+)/GLUCOSE COTRANSPORTER
 ARRANGED IN SERIES ALONG KIDNEY PROXIMAL TUBULES.
 -1- SUBCELLULAR LOCATION: Integral membrane protein.
 -1- DEVELOPMENTAL STAGE: APPEARS ON EMBRYONIC DAY 18 AND GRADUALLY
 INCREASES UNTIL BIRTH.
 -1- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; U03120; AAA19015.1; -;
 EMBL; D16101; BAA03676.1; -;
 EMBL; AB000729; BAA19172.1; -;
 InterPro: IPR001734; Na/solut_symport.
 Pfam: PF00474; SSF; 1.
 TIGRfam: TIGR00813; sss; 1.
 PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
 PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
 PROSITE; PS00283; NA_SOLUT_SYMP_3; 1.
 Transport; Sugar transport; Transmembrane; Sodium transport; Symport;
 Glycoprotein.
 KW DOMAIN 1 28 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 29 47
 FT DOMAIN 48 64 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 65 85
 FT TRANSMEM 86 105 POTENTIAL.
 FT TRANSMEM 106 126 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 127 171 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 172 191 POTENTIAL.
 FT TRANSMEM 193 208 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 209 229 POTENTIAL.
 FT TRANSMEM 230 270 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 271 291 POTENTIAL.
 FT TRANSMEM 292 314 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 315 334 POTENTIAL.
 FT TRANSMEM 335 423 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN

TRANSMEM 424 443 POTENTIAL.
 FT DOMAIN 444 455 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 456 476 POTENTIAL.
 FT DOMAIN 477 526 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 527 547 POTENTIAL.
 FT TRANSMEM 548 643 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 644 664 POTENTIAL.
 FT CARBOHYD 248 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 354 Y -> H (IN REF. 3).
 SQ SEQUENCE 665 AA; 73066 MW; A92038D964BFF061 CRC64;
 Query Match 10.3%; Score 306; DB 1; Length 665;
 Best Local Similarity 23.5%; Pred. No. 7.4e-13;
 Matches 155; Conservative 105; Mismatches 242; Indels 158; Gaps 29;
 QY 11 ITVFYLLILLVIGWAAWTKNSGSAERSEALIVGGRDGLLVGGFTMTATWVGYYING 70
 DB 32 IVIYFVWVAVGLWAMFST-NRGTV-----GGFFLAGRSMWVWPGISAFASNGSHFVG 86
 QY 71 TAEAVYVPGYGLAWAQAIPIGYSLS-----LILGGLFFFAKPMRSK-GYVTMLDPFQOYIK 124
 DB 87 LA-----GTGAAGATMGGFENALVFVVLGWLFFV--PIYIKAGVVTM-----PEYLRK 134
 QY 125 RMGG-----LFFPALMGEMFWAAAFSALGATISVIIDVDHHSIVISALIA 172
 DB 135 RFGKRIQIVLSVLSLLYIFTKISADIFSGAIF-----INLAUGLDIYLAIFILLAIT 188
 QY 173 TLTLYGGLYSVAYTVVOLFCIFVGLWTSVPEALSHPAVADIGFTAVHAKYOK--PWL- 229
 DB 189 ALYITIGGGLAAVYITDTLQTAIMLVGSFILTGFAPREVG---GYEAFMDKYMKAIPTLV 244
 QY 230 --GTVD--SSEVYS-WLDSFLL--MLGGIPW-----OAYFORVLS 264
 DB 245 SDGNITVKECYTPRADSHFIFRDPITGDMPPGLIFGLSILALWYCTDQVIVQCLSA 304
 QY 265 SSATYAOVLSFAAGCLVMAIPAILGAGTSDMNTAYGLPDP-----KTTEADM 318
 DB 305 KNMSHVAGCTLCGLKLLPMLFVMPGMSRLIYTDKIACVLPSECKKYCGTPVGCTNI 364
 QY 319 ILPIVLYVCPVYISFFGLGAVSAVMSADSSILSSAFARNIYQLSFQNASDKIEV 378
 DB 365 AYPILVVELMPNGLRGLMLSVMAASLSSLTSEFNSASTLTFTMDIY--TKIRKGASEKELM 423
 QY 379 WVMRITVEV-FCASATAMALTKTVYG--LWYLSDDLVI--VIFPQLLCVLFVKGNTRY 433
 DB 424 IAGRLFILVLIGISIAWVPIVQSSAQSGQLFDYIQISYLSYLPPIAAVFAIFCKRVNEP 483
 QY 434 GAVAGYVSGFLRI-----TG-----GEPYLYLQPLIFPGYYPDDN 470
 DB 484 GAFWGLILGLIGISRMITEFAYGTGSCMEFSCNPKIICGVHYLYFAILF----- 534
 QY 471 GIYNOKFPFKLTAMVTSFLTNIICISYLAKYLPESGTLPPKLDVDAV--VARHSEENMDK 528
 DB 535 -----AISVVTVLIVISLTKPI-----PDVHLVRLWLSLRNSTEERID- 572
 QY 529 TILVKNENIKLDELALVKPQSMSTSSFTNKE-----AFLVDSSPESGTGD 577
 DB 573 --LDAGEEPVEE---DPKDTIEIDAEAPQKEGCKFRKAYDLFCGLQDQKPKMKKEE 626
 RESULT 5
 SL54_PIG STANDARD; PRT; 660 AA.
 ID SL54_PIG
 AC P31636;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Low affinity sodium-glucose cotransporter (Sodium/glucose
 cotransporter 3) (Na(+)/glucose cotransporter 3).
 GN SLC5A4 OR SGLT3 OR SGLT1.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

CC -!- TISSUE SPECIFICITY: MORE ABUNDANT IN HEART THAN IN KIDNEY, WHERE
CC IT IS ABSENT FROM THE OUTER CORTEX.
CC -!- SIMILARITY: BELONGS TO THE SODIUM: Solute symporter family (SSF).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC EMBL; M84020; AAA31421.1; -
CC PIR; A42251; A42251.
CC InterPro; IPR001734; Na/solut_symport.
CC Pfam; PF00474; SSF; 1.
CC TIGRFAMs; TIGR00813; sss; 1.
CC PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
CC PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
CC PROSITE; PS00283; NA_SOLUT_SYMP_3; 1.
CC Transprot; transmembrane; Sodium transporter; Symport; Glycoprotein.
CC DOMAIN 1 25
CC CYTOPLASMIC (POTENTIAL).
CC POTENTIAL.
CC 26 44
CC EXTRACELLULAR (POTENTIAL).
CC 45 61
CC POTENTIAL.
CC 62 82
CC CYTOPLASMIC (POTENTIAL).
CC 83 102
CC POTENTIAL.
CC 103 123
CC EXTRACELLULAR (POTENTIAL).
CC 124 168
CC POTENTIAL.
CC 169 188
CC CYTOPLASMIC (POTENTIAL).
CC 189 205
CC POTENTIAL.
CC 206 226
CC EXTRACELLULAR (POTENTIAL).
CC 227 270
CC POTENTIAL.
CC 271 291
CC CYTOPLASMIC (POTENTIAL).
CC 292 314
CC POTENTIAL.
CC 315 334
CC EXTRACELLULAR (POTENTIAL).
CC 335 422
CC POTENTIAL.
CC 423 442
CC CYTOPLASMIC (POTENTIAL).
CC 443 454
CC POTENTIAL.
CC 455 475
CC EXTRACELLULAR (POTENTIAL).
CC 476 525
CC POTENTIAL.
CC 526 546
CC CYTOPLASMIC (POTENTIAL).
CC 547 650
CC POTENTIAL.
CC 651 671
CC N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 250
CC SEQUENCE 672 AA; 73161 MW; E2D987B03B9C57B4 CRC64;

Query Match 10.0%; Score 298; DB 1; Length 672;
Best Local Similarity 25.0%; Pred. No. 2.5e-12;
Matches 153; Conservative 89; Mismatches 232; Indels 138; Gaps 25;
9 IATV-VFYLILVGLVGLSMCRT-NRGTV-----GGYFLAGRWVWVPGVSLFASNIGSGH 67
26 IAVIAAYFLVIGVGLSMCRT-NRGTV-----GGYFLAGRWVWVPGVSLFASNIGSGH 80
68 INGTAFAVYVPGYGLAQAQPIGYSLS-----LIIGGLFFAKPMKSKYVVMPLDPFQIYG 123
81 FVGLA-----GTGAAGLAVAGFENWALFVLLGLWFLFAPVLTAGVITM-----POYL 130
124 KRMGG-----LLFPLMGEMFWAAAF-----SALGATISVIIDVDMHISVIISA 169
131 KRFGGHRIRLYSLVSLFTYFTKISVDVDFGAFVIOALGNW-----YASVIAL 182
170 LIATLYLVGLVSVAYTDVQVLCIFVGLVSWVPFALSHAPADIGTAVHAKY----- 224
183 GITWYVTVTGLAALMYTDVQVLCIFVGLVSWVPFALSHAPADIGTAVHAKY----- 238
225 -----QKPLGTVDSVYVWLDSPFLL-----MLGIPW-----QAYF 258
239 SLTVSDEDPAGVNISSCYRPRDSYHLLRDFVTGDLPPALLGLTIVSGVWMSQDIV 298
259 QRVLSSTATYVQVLSFAAFCGLVMAIPAILIGAGASTDMNQATYGLDPKPT-----TE 314
299 QRLAGRLNTHIKACILCGYLLKLPFPMFLMWPMGMSIRLYPDEACVAPKCVKRGVTE 358

QY 315 E--ADMILPTVLOVCPVYISFFGLGAVSAVNVSSADSSILSASSMFARNIYQLSERQNA 372
Db 359 VGCSSVPRLVKVLMPENGLRGLMAYLAALMSSLSIFNSSSTLTMDIYTL--RPRA 416
QY 373 SDKEIVWVRITVFVFCASATAMALLTKTVG---LWVLSDDLVIIV--IPFQLLCVLFV 427
Db 417 GEGELLVGRUWVFIIVAVSVAWLFPVQAAGGOLFDFYIQSVSSVLAAPPVSAVFVALFV 476
QY 428 KGTNYGAVAGYVSGFLRITGGEPYLYLQPLIFYPGYPDNGIYNGKFPFKTLAMV-- 485
Db 477 PRVNEKGAFWGLGGLMLGLARLP-----EFSGFGSCVRP 513
QY 486 ---TSFLTNICISYLAFLFE-SG-----TLP-PKLDVDFDAVVA-RISEENMDKTI 530
Db 514 SACPAFLCRVHYLFAIVLFFCSCGLLIIVSLCTAPIPRKHHLRVLFSLRHSKE----- 567
QY 531 LVKNEIKLDEL 542
Db 568 --EREDLDADEL 577
RESULT 7
ID SL54 HUMAN STANDARD; PRT; 659 AA.
AC Q9NV91; O15279;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Low affinity sodium-glucose cotransporter (Sodium/glucose
DE cotransporter 3) (Na(+)/glucose cotransporter 3).
GN SLC5A4 OR SAAAT1 OR SGLT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Small intestine;
RT "The molecular cloning and functional characterization of the human
RT SGLT2 transporter";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
RA Clump M., Skink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burdill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.C.T.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon L., Stewart C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Aoki N., Mitsuyama S.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hua P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.P., Loh P., Malaj E., Nguyen T., Pan H.,

QY 177 LVGGLYSVAYTDVYVQVLCIFVGLWISVPFALSHPAVADIGTAVHAKYQKPLGTVD--- 233
Db 193 ITGGLASVIYTDVQAVIMLVGSFILMVAF---VEVGGYSFTEKFMNAIPSVVEGDN 248
QY 234 ---SSRVS-WLDSFLL---MLGGIPW-----QAVFQVRLSSSAT 268
Db 249 LTNISRCYTPQDSHFIRDPVTGIPWPGTAFGMPITALWYWCINQVIVQRLCGKNLS 308
QY 269 YAVLSFLAFAAGLWMAIPAILGAIGASTDMNQAYGLP-----DPKTEEADMI 319
Db 309 HVKACILCGYKLLPLFFWMPGMSIRLITDMWACVVPSECVKHCVDVGCTNYA--- 365
QY 320 LPVLQYLCPPYISFFGLGAVSAVMSADSSILSSAFNRIYQLSFRONASDKETVM 379
Db 366 YPMVLKMPGLRGLMLSVMLASLMSSTSVFNSASTLETDLY-TKIRKASERELLI 424
QY 380 VMRTVFVFGASATAMALLTKTVYG---LWLSSDLVYI--VIPOLLCLVLFVGTNTYG 434
Db 425 AGRFVSVLIVTSILKWPIVEVSGGQLVHYTEALSSYLGPPIAAVFLVAVFCKRANEQ 484
435 AVAGVSGSLF-----RITGGEPLYLQPLIFYPGYPPDDNG 471
Db 485 AFWGLWGLVGLLRLMIAEFSYGRGSLAPSSCPKIICGVHYLYFAILFF----- 535
QY 472 IYNOKFPKTLAMVTSFLTNICISYLAFLFESGLPKLDV---PDVAVHSEENMD 527
Db 536 -----VCILVILGSYLTK-----PIPDVHLHRLCWALRNSKEERID 572
QY 528 KTIIVKNEN--IKLELALVKPR-----QSMTSLSTFTNKAEFLVDVSSP 570
Db 573 LDAREENGADRTBEDQTEPRGLKKTCDLFCGLQRAEFKLTKEVEEALDTTEKP 631

RESULT 12
SL53_MOUSE
ID SL53_MOUSE STANDARD; PRT; 718 AA.
AC Q9JKZ2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN Sodium/myo-inositol cotransporter (Na(+)/myo-inositol cotransporter).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP
RX MEDLINE=2037552; PubMed=10773690;
RA McVeigh K.E., Mallee J.J., Lucente A., Barnoski B.L., Wu S.,
RT Berry G.T.;
RT "Murine chromosome 16 telomeric region, homologous with human
cotransporter (SLC5A3) gene.";
RL Cytogenet. Cell Genet. 88:153-158(2000).
CC -1- FUNCTION: PREVENTS INTRACELLULAR ACCUMULATION OF HIGH
CONCENTRATIONS OF MYO-INOSITOL (AN OSMOLYTE) THAT RESULT IN
IMPAIRMENT OF CELLULAR FUNCTION.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AF220915; AAF43668.1;
CC MGD; MGI:1858226; SLC5a3.
CC InterPro; IPR001734; Na/solut_symport.
CC Pfam; PF00474; SSF; 1.

TIGRFAMS; TIGR00813; sss; 1.
DR PROSITE; PS00456; NA_SOLUT_SYMP_1; 1.
DR PROSITE; PS00457; NA_SOLUT_SYMP_2; 1.
DR PROSITE; PS0283; NA_SOLUT_SYMP_3; 1.
KW Transport; Transmembrane; Sodium transport; Symport; Glycoprotein.
FT DOMAIN 1 9
FT TRANSMEM 10 29
FT DOMAIN 30 38
FT TRANSMEM 39 57
FT DOMAIN 58 86
FT TRANSMEM 87 110
FT DOMAIN 111 123
FT TRANSMEM 124 144
FT DOMAIN 145 157
FT TRANSMEM 158 183
FT DOMAIN 184 186
FT TRANSMEM 187 205
FT DOMAIN 206 303
FT TRANSMEM 304 324
FT DOMAIN 325 353
FT TRANSMEM 354 376
FT DOMAIN 377 406
FT TRANSMEM 407 430
FT DOMAIN 431 443
FT TRANSMEM 444 462
FT DOMAIN 463 510
FT TRANSMEM 511 532
FT DOMAIN 533 695
FT TRANSMEM 696 716
FT CARBOHYD 32 32
FT SITE 24 24
FT SITE 285 285
SQ SEQUENCE 718 AA; 79554 MW; D035CFBECDDA803B CRC64;

Query Match 9.7%; Score 289; DB 1; Length 718;
Best Local Similarity 21.7%; Pred. No. 1e-11;
Matches 150; Conservative 113; Mismatches 209; Indels 218; Gaps 32;
QY 9 IALL-VEYLLILLVGIWAARFKNSGAESEAIIVGGRDILGLVGGTMTATWYGGG- 66
Db 10 IAAVALYFLVMCIIGFFAMKSNRSTVS-----GYFLAGRSMTWAIGA 53
QY 67 ---YINGTAEAVYVPGYGAQAAPIGYS-----LSLILGLGLFFAKPMRSKGYVTM 114
Db 54 SLFVSNIGSEHFI---GLAGSAAAGFAVAGWAFNALLQLLGLWVFIPIYRS-GVYTM 109
QY 115 LDPFQIYGRKMG-----LLFIPALMGEMFWAAAFSALGATISVIIDVDMH 162
Db 110 ---PEYLSKRFGRHIOVYFAALSLLYIFTKLSVDLYSGALF-----IQESLGNWLY 159
QY 163 ISVIISALIAITLYTLVGLGYVAVYDVVQLFCIFVG-----LWISV-----PPAL 207
Db 160 VSVILLIGMTALLVTGGLVAVIYTDLOALLMIIGALTLMVISMKVKGFEVKRRYML 219
QY 208 SHPAVADI-----GFTAVHAKYQK-----PWLGTV---DSSEYVSWLDS 243
Db 220 ASPDVASILLKYNLSNTNACMVHPKANALKMLRDPDTEDEVPWPGFILQTPASVWYWCAD 279
QY 244 FLLMLGGIPQWQAYFORVLSSSSATYAO-----VLSFLAAGCLVMAIPAIL----- 290
Db 280 -----QVIVQVRLAAKNIAHAKGSLTMAGFLKLLPFMTIVYVPGMISRVFADEI 328
QY 291 -----IGAIGASTDMNQAYGLPDKPTEEDMILPIVLQVLCPPYISFFGLGAVSA 342
Db 329 ACINPEHCQVCGSAGCSNIAI-----PRLVMTLPVPGURLMMAVMA 373
QY 343 AVMSADSSILSSAFNRIYQLSFRONASDKETIVVWVRITV-FVFGASATAMALLTKT 401
Db 374 ALMSDLDLSTFNASTIFTLDVYKL-IRKSASSRELIMVGRIFVAFVWVVISVWPIIVEM 432
QY 402 VYGLWYLSDDLVIYVIFPQL-----LCVLFVKGTNT-----YGAVAGYVSG---LFLRITGG 450

Query Match	9.68;	Score 285;	DB 1;	Length 452,
Best Local Similarity	22.18;	Pred. No. 1.2e-11;	Models 106;	Gaps 18;

Query Match	9.5%; Score 2.75; Pred. No. 8.5e-11;	Indels	172;	Gaps	32;
Best Local Similarity	22.2%;	Mismatches	225;		
Matches	148; Conservative	122;			

QY	571	EGSGTED	577
	:		
Ph	505	KGIOPED	602

Search completed: March 31, 2003
Job time : 42 secs

```

SQ SEQUENCE 718 AA; 79673 MW; 206BE25FA385111D CRC04;

```


GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 18:28:26 : Search time 49 Seconds
(without alignments)
1137.918 Million cell updates/sec

Title: US-10-069-541-6
Perfect score: 2972
Sequence: 1 MAFHVEGLIAIIVFYLLIL.....EAFLDVDSPEGSGETDNLQ 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2972	100.0	580	JC7502	choline transporter
2	1361.5	45.8	631	T20037	hypothetical prote
3	344	11.6	492	D75188	proline symporter
4	308.5	10.4	662	A37226	glucose transport
5	306	10.3	664	A33545	Na+/glucose cotran
6	303.5	10.2	665	A53582	Na+/glucose cotran
7	303.5	10.2	660	A44432	amino acid transpo
8	301	10.1	463	E83468	probable sodium/so
9	299.5	10.1	507	B83988	proline transporte
10	298	10.0	672	A42251	nucleoside transpo
11	294	9.9	664	S59637	glucose transport
12	292	9.8	672	A56765	sodium-glucose cot
13	288	9.7	664	S59638	glucose transport
14	286	9.6	491	H71097	hypothetical prote
15	285	9.6	492	D69670	sodium/proline sym
16	284	9.6	537	C71008	probable proline p
17	283	9.5	484	E75138	osmoregulated prol
18	282.5	9.5	501	T44298	sodium/proline sym
19	279.5	9.4	496	H82382	sodium/proline sym
20	278.5	9.4	718	A42163	Na+/myo-inositol c
21	271	9.1	537	A75123	proline permease (
22	270.5	9.1	718	A56851	Na+/myo-inositol c
23	267	9.0	522	B84247	proline permease f
24	263.5	8.9	494	JC2382	sodium/proline sym
25	261	8.8	512	E89978	high affinity prol
26	260.5	8.8	504	E64118	sodium/proline sym
27	260.5	8.8	605	A36361	glucose transport
28	259	8.7	526	C69115	sodium/proline sym
29	254	8.5	496	A71980	sodium/proline sym

ALIGNMENTS

RESULT 1

JC7502

choline transporter - human

C;Species: Homo sapiens (man)

C;Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 01-Dec-2000

C;Accession: JC7502

R;Apparundaram, S.; Ferguson, S.M.; George Jr., A.L.; Blakely, R.D.

Biochem. Biophys. Res. Commun. 276, 862-867, 2000

A;Title: Molecular cloning of a human, hemicholinium-3-sensitive choline transporter

A;Reference number: JC7502

A;Contents: Spinal cord

A;Accession: JC7502

A;Molecule type: mRNA

A;Residues: 1-580 <APP>

A;Cross-references: DB:AF276871

C;Comment: This protein, a hemicholinium-3-sensitive phosphorylated transmembrane p

C;Genetics:

A;Gene: cht

A;Map position: 2q12

C;Keywords: choline transport; spinal cord; transmembrane protein; transport protei

Query Match 100.0%; Score 2972; DB 2; Length 580;
Best Local Similarity 100.0%; Pred. No. 8.2e-211; Indels 0; Gaps 0;
Matches 580; Conservative. 0; Mismatches 0;

Qy	1	MAFHVEGLIAIIVFYLLILLVGIWAARTKNSGSAERSEAIIVGGRDIGLLVGGFTMTA	60
Db	1	MAFHVEGLIAIIVFYLLILLVGIWAARTKNSGSAERSEAIIVGGRDIGLLVGGFTMTA	60
Qy	61	TWVGGYINGTAEAVVPGYGLAWAQAPIGYSLSLILGLFFAKPMRSKGYVTMLDPFQQ	120
Db	61	TWVGGYINGTAEAVVPGYGLAWAQAPIGYSLSLILGLFFAKPMRSKGYVTMLDPFQQ	120
Qy	121	IYKRMGGLLFTPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIALTYLVGG	180
Db	121	IYKRMGGLLFTPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIALTYLVGG	180
Qy	181	LYSVATDVVQLFCIFVGLWISVPFALSHPVADIGFTAVHAKYQKPMGLTVDSSEYVS	240
Db	181	LYSVATDVVQLFCIFVGLWISVPFALSHPVADIGFTAVHAKYQKPMGLTVDSSEYVS	240
Qy	241	LDSFLLMLGGIPWQAFYFQVLSSSATYAQVLSFLAAGCLVMAIPAILIGAISTDW	300
Db	241	LDSFLLMLGGIPWQAFYFQVLSSSATYAQVLSFLAAGCLVMAIPAILIGAISTDW	300
Qy	301	NOTAYGLPDPKTEEDMILPIVLYLCPVYISFFGLGAVSAAVMSADSILLSASSMFA	360
Db	301	NOTAYGLPDPKTEEDMILPIVLYLCPVYISFFGLGAVSAAVMSADSILLSASSMFA	360
Qy	361	RNIYQLSFRONASDKELVWVRITVFVGASATAMALLTKTVYGLWYLSLDLYVIVFPQ	420
Db	361	RNIYQLSFRONASDKELVWVRITVFVGASATAMALLTKTVYGLWYLSLDLYVIVFPQ	420

30	254	8.5	502	2	E90786	major sodium/proli
31	254	8.5	502	2	C85646	major sodium/proli
32	253.5	8.5	497	2	T48676	proline uptake pro
33	253	8.5	502	1	JGECPP	sodium/proline sym
34	252	8.5	461	2	G83610	probable sodium/so
35	249	8.4	496	2	G64526	sodium/proline sym
36	249	8.4	502	2	AF0633	sodium/proline sym
37	246.5	8.3	512	2	S75887	hypothetical prote
38	246	8.3	502	2	S10220	sodium/proline sym
39	243.5	8.2	506	2	F83547	sodium/solute symp
40	243	8.2	497	2	AH2015	sodium/proline sym
41	242	8.1	449	2	B69759	pantothenate trans
42	239.5	8.1	484	2	H64105	pantothenate trans
43	239	8.0	549	2	B65215	hypothetical 59.2
44	237.5	8.0	577	2	T28017	hypothetical prote
45	236	7.9	491	2	E69383	pantothenate perme

QY 421 LLCVLEVKGTNTYGAVAGYVSGFLRLRTGGEPELYLQPLIFPGYYPDDNGIYNQKPEFK 480
||||| : : : : :
Db 421 LLCVLEVKGTNTYGAVAGYVSGFLRLRTGGEPELYLQPLIFPGYYPDDNGIYNQKPEFK 480
||||| : : : : :
QY 481 TLAMVTSFLTNICISYLAKYLFPSGTLPPKLDVDFVAVARHSEENMDKTLVKNENIKLD 540
||||| : : : : :
Db 481 TLAMVTSFLTNICISYLAKYLFPSGTLPPKLDVDFVAVARHSEENMDKTLVKNENIKLD 540
||||| : : : : :
QY 541 ELALVKPROSMTLSSTFTNKEAFLDVDSPEGSGETEDNLQ 580
||||| : : : : :
Db 541 ELALVKPROSMTLSSTFTNKEAFLDVDSPEGSGETEDNLQ 580
||||| : : : : :
RESULT 2
T20037
hypothetical protein C48D1.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20037
Burton, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19214
A:Accession: T20037
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-631 <SWIL>
A:Cross-references: EMBL:Z81049; PIDN: CAB02847.1; GSPDB: GN00022; CESP: C48D1.3
A:Experimental source: clone C48D1
C:Genetics:
A:Gene: CESP:C48D1.3
A:Map position: 4
A:Introns: 82/1; 120/3; 187/1; 236/3; 249/1; 358/1; 510/3; 570/2

Query Match 45.8%; Score 1361.5; DB 2; Length 631;
Best Local Similarity 46.8%; Pred. No. 2.3e-92;
Matches 290; Conservative 91; Mismatches 146; Indels 93; Gaps 10;
QY 7 GLIAIVFYLLVGVINWARTKNSGAER-----SEAIIVGGRDIGLLVGGFTMTATW 62
||| : : : : :
Db 6 GIVAIIFYVLLVGVINWARTKNSGAER-----SEAIIVGGRDIGLLVGGFTMTATW 65
||| : : : : :
QY 63 VGGYINGTAEAVVPGYGLAWAQAPIGYSLSLILGGLFFAKPMRSKGVVTMLDP----- 117
||||| : : : : :
Db 66 VGGYINGTAEALY--NGLLGCGAPVGYAISLVNGLLFLAKMKEEYITMLDPFQFWN 123
||||| : : : : :
QY 118 -----FQYIGKRMGGLLFPALMGEMFWAAATF 146
||| : : : : :
Db 124 FLEIFGRTFDFNFKLGRFLKLTIIILDFPFQHKYQRIIGLMVVPALLGETFWTAATL 183
||| : : : : :
QY 147 SALGATISVIIDVDMHISVILSALITATLYLVGGYLVAYTDVQLFCIFVGLWISVPPA 206
||||| : : : : :
Db 184 SALGATLSVILGIDMNASVILSACIAVFTTGGYIAYTDVQLFCIFVGLLILGLYV 243
||||| : : : : :
QY 207 LSHPAVADIGTAVHAKYQKPMWLTVDSSSEYSLWLSFLLMLGGIPWQAYFQVLSSSS 266
||| : : : : :
Db 244 QNRPN-----REFKSLWIDCMLLVFGGIPWQVYFQVLSST 282
||| : : : : :
QY 267 ATVAQVLSFLAAFGCLVMAIPAILIGAGTDNQTAYGLDPDKTTEA-----DMIL 320
||| : : : : :
Db 283 AHGQATLSFVAGVGCILMAIPALIGATARNDRMTDYSPPNNGTKVESIPPPKRNMY 342
||| : : : : :
QY 321 PIVLQYLCPVYISFPLGAVSAVWSSADSLSSASSFARNIYOLSPRONASDKEIYVW 380
||||| : : : : :
Db 343 PLVFOYLTFRWAFVIGLAVSAVWSSADSLSSASSFARNIYOLSPRONASDKEIYVW 402
||||| : : : : :
QY 381 MRITVFEVASATAMALLTITVGLWLSDDLVIIVITPQLLCVLFVKGNTNYGAVGYV 440
||||| : : : : :
Db 403 MRATICVGMATIMATIOSIYGLWLCADLVVILFPQLLCVVMYPRNSNYGSLAGYA 462
||||| : : : : :
QY 441 SGLFRITGGEPELYLQPLIFPGYYPDDNGIYNQKPEKTLAMVTSFLTNICISYLAKY 500
||||| : : : : :
Db 463 VGLVRLIGGEPLVSLPAFFHYPMY---TGCV--QYFFRTTAMLSMATIIVYSIOSEK 517
||||| : : : : :

QY 501 LFESTLPPKLDVDFVAVARHSEENMDKTLVKNENIKLDELALVKPROSMTLSSTFTNK 560
||||| : : : : :
Db 518 LFKSRLSPEDVMGCV-----NIPIDHVLPSD-VSFAYSSETLNM 559
||||| : : : : :
QY 561 EAFLDVDSPEGSGETEDNLQ 580
||||| : : : : :
Db 560 KVECDGMQFPQ-LQTEHRLQ 578
||||| : : : : :
RESULT 3
D75188
proline symporter (proline permease). PAB2354 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: D75188
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome
A:Reference number: A75001
A:Accession: D75188
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-492 <KAW>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB48955.1; PID:g5457433
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: putP-3; PAB2354
C:Superfamily: proline carrier protein

Query Match 11.6%; Score 344; DB 2; Length 492;
Best Local Similarity 24.2%; Pred. No. 1.1e-17;
Matches 132; Conservative 99; Mismatches 196; Indels 118; Gaps 25;
QY 8 LIAIIVFYLLVGVINWARTKNSGAERSEAIIVGGRDIGLLVGGFTMTATWVGGY 67
||| : : : : :
Db 14 LVAFLLTLLPILVGFYAMKRTK-----EEDFFVGGGRAMDKITVALSAVSSGRSSWL 66
||||| : : : : :
QY 68 INGTAEEAVVPGYGLAWAQAPIGYSLS-----LILGGLFFAKPMRSKGVVTMLDPQQIYG 123
||| : : : : :
Db 67 VLGLSGWAYKMGVTAVW--RAVGVIVAEFMFVYNGIRLKFSEFNATVPDYEARPR 124
||||| : : : : :
QY 124 K-----RMGG-----LLFIPALMGEMFWAAATFSAIGATISVIIDVDMHISVILSALITL 174
||| : : : : :
Db 125 DTSKILRIAASIIIIIFLTSVGAQFNAGA-----KTLSTALGISIETLALMISVLMIIV 178
||||| : : : : :
QY 175 YTLVGGYLVAYTDVQLFCIFVGLWISVPELALSHPAVADIGFT-----AVHAKYQK 226
||| : : : : :
Db 179 YMLGGFIAYANDVIRAVIMIIGLVV-----LPVIAVAKVGGTEEVLVKVLHALDPKLN 233
||| : : : : :
QY 227 PW---LGTVDSSSEYSLWLSFLLMLGL-GIPWQAY-FQVLSSSSSATVAQVLSFLAAFGC 281
||| : : : : :
Db 234 PMAFAGVVG-----FLGIGFGSPGPHIIVRYMSIDDPNKLRYSTVVGTWN 282
||||| : : : : :
QY 282 LVMAIPAILIGAGTDNQTAYGLDPDKT---EEDMILP-IVLQVLCVYISFVGLG 338
||| : : : : :
Db 283 VILAWGAIFFVLAGRAI-----VPDVSQLPKNAEMIPYLSAQYFPPLILYGL-IG 333
||| : : : : :
QY 339 AVSAAVWSSADSLSSASSFARNIYOLSPRONA--SDKEIVWVRITVFEVASATAMA 396
||||| : : : : :
Db 334 GIFAAILSTADSQLVAVSTVVKDLQEVILKGTIDEKALTISRVTLVVGLAALIA 393
||||| : : : : :
QY 397 LUTKTVYGLWYSSDLVY-IVIF-----PQLCVLFVKGNTNYGAVGYVSGLFL 445
||||| : : : : :
Db 394 -----YAKDIIFFVLFVFWAGGLGASFGPTLILSLYKGTWKWGLAGMIVGIT 443
||||| : : : : :
QY 446 RTTGGEPELYLQPLIFPGYYPDDNGIYNQKFPKTLAMVTSFLTNICISYLAKYLESG 505
||||| : : : : :
Db 444 TIVW---KYLKPI-----TGLY-ELVP-----AFIFSUIATIIVSMITK----- 479
||| : : : : :
QY 506 TLPPK 510
||| : : : : :
Db 480 --PPE 482

QY 437 AGYVSGLELRI-----TG-----GEPYLYLOPLIFPGYPPDNGIY 473
 Db 487 WGLVGLFGLIGSRMITEFAYCTGSCMEPSNCPTIICGVHLYLYFAILL- 534
 QY 474 NQKPFKTLAMVTSFLTNICISYLAFLFSGTLPKLDVFDVVA-RHSEBMDKTLIV 532
 Db 535 -----VISIITVVVSLFTKPI-----PDVHLRLCWSLRNSKE----- 568
 QY 533 KNEIKLD-ELALVKPQSMTSLSTFTNKEAF-----LDVDSSEPGSGTDED 577
 Db 569 --ERIDLDAGEEDIOEAPEEATDEVPKKKGFFRAYDLFCGLDODKPKMTKEE 623

RESULT 5

A33545
 Na+/glucose cotransporter SGLT1 - human
 C:Species: Homo sapiens (man)
 C:Date: 27-Feb-1990 #sequence_revision 27-Feb-1990 #text_change 20-Aug-1999
 C:Accession: A33545; A53804
 R:Hediger, M.A.; Turk, E.; Wright, E.M.
 Proc. Natl. Acad. Sci. U.S.A. 86, 5748-5752, 1989
 A:Title: Homology of the human intestinal Na(+)/glucose and Escherichia coli Na(+)/P
 A:Reference number: A33545; MUID:89345544; PMID:2490366
 A:Accession: A33545
 A:Molecule type: mRNA
 A:Residues: 1-664 <HED>
 A:Cross-references: GB:M24847; NID:G338054; PIDN:AAA60320.1; PID:G338055
 R:Turk, E.; Martin, M.G.; Wright, E.M.
 J. Biol. Chem. 269, 15204-15209, 1994
 A:Title: Structure of the human Na+/glucose cotransporter gene SGLT1.
 A:Reference number: A53804; MUID:94253082; PMID:8195156
 A:Accession: A53804
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-45 <TUR>
 A:Note: sequence extracted from NCBI backbone (NCBIN:147993, NCBI:147994)
 C:Genetics:
 A:Gene: GDB:SLC5A1; SGLT1
 A:Cross-references: GDB:I120375; OMIM:182380
 A:Map position: 22q13.1-22q13.1
 C:Superfamily: proline carrier protein
 C:Keywords: transmembrane protein; transport protein

Query Match 10.3% Score 306; DB 2; Length 664;
 Best Local Similarity 22.8%; Pred. No. 9.9e-15;
 Matches 148; Conservative 104; Mismatches 218; Indels 178; Gaps 30;

QY 11 IIVFYLLLLVGIWAARTKNSGAERSEAIIVGGRDIGLLVGGFTWTATVWGGYING 70
 Db 32 IIVFYVVMVAVGLWAMEST-NRGTV-----GGFFLAGRSMVWMPIGASLFASNIGSGHFGV 86
 QY 71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSK-GYVTMLDPFQOIYK 124
 Db 87 LA-----GTGAASGIATGGFEWNNALVVLVGLWLFV--PIYIKAGVVTM-----PEYLRK 134
 QY 125 RMGG-----LLFIPALMGEMFWAAAFSALGATISVIIDVDMHISVIISALIA 172
 Db 135 RFGGRIQVYLSLLSLLLYFTKISADIFSGAIF-----INLALGLNLYLAIFLLAIT 188
 QY 173 TLYTLVGGLYSVAYTDVQLFCIFVGLWISVPFALSHPAVDIGTAVHAKYOK--PWL- 229
 Db 189 ALYTTITGGAAVYITDTLTQVIMLVGSLILTGFAPFHEVG-----GYDAFPEKYNKAFTIV 244
 QY 230 ---GIVDSSEVYS-WLDSFLL--MLGGIPW-----QAYFORVLSS 264
 Db 245 SDGNTTFQEKCYTPRADSFHIFRDPLTGDLFPWGFIFGMSILTLWYCTQDQIVQRCLSA 304
 QY 265 SSATYAQ-----VLSFLAAFGCLVMAIPAIL-----TGAI-----GASTDWNQT 303
 Db 305 KNSHVKGCGCILGKYLKMPMTWMPGMISRLIYTEKIACTVVPSECEKCYGKVGCTNI 364
 QY 304 AYGLPDPKTTTEADMLPIVLYOYLCPYVIFFGGLGAVSAAMSSADSILLSASSMFARNI 363

RESULT 4

A37226
 Na+/glucose cotransporter - rabbit
 N:Alternate names: sodium/D-glucose cotransporter
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 30-Dec-1991 #sequence_revision 01-Mar-1996 #text_change 20-Aug-1999
 C:Accession: S00515; S15974; A37226
 R:Hediger, M.A.; Coady, M.J.; Ikeda, T.S.; Wright, E.M.
 Nature 330, 379-381, 1987
 A:Title: Expression cloning and cDNA sequencing of the Na+/glucose co-transporter.
 A:Reference number: S00515; MUID:88065856; PMID:2446136
 A:Accession: S00515
 A:Molecule type: mRNA
 A:Residues: 1-662 <HED>
 A:Cross-references: EMBL:X06419; NID:g1640; PIDN:CAA29727.1; PID:g1641
 R:Morrison, A.I.; Panayotova-Heller, M.; Feigl, G.; Schoelermann, B.; Kinne, R.K.H.
 Biochim. Biophys. Acta 1089, 121-123, 1991
 A:Title: Sequence comparison of the sodium-D-glucose cotransport systems in rabbit renal
 A:Reference number: S15974; MUID:91223090; PMID:2025641
 A:Accession: S15974
 A:Molecule type: mRNA
 A:Residues: 1-662 <MOR>
 A:Cross-references: EMBL:X55355; NID:g1716; PIDN:CAA39040.1; PID:g1717
 R:Coady, M.J.; Pajor, A.M.; Wright, E.M.
 Am. J. Physiol. 259, C605-C610, 1990
 A:Title: Sequence homologies among intestinal and renal Na(+)/glucose cotransporters.
 A:Reference number: A37226; MUID:91023017; PMID:2221040
 A:Accession: A37226
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 178-662 <COA>
 A:Cross-references: GB:X06419
 A:Experimental source: renal cortex
 C:Superfamily: proline carrier protein

Query Match 10.4% Score 308.5; DB 2; Length 662;
 Best Local Similarity 23.4%; Pred. No. 6.5e-15;
 Matches 154; Conservative 110; Mismatches 238; Indels 155; Gaps 26;

QY 11 IIVFYLLLLVGIWAARTKNSGAERSEAIIVGGRDIGLLVGGFTWTATVWGGYING 70
 Db 32 IIVFYVVMVAVGLWAMEST-NRGTV-----GGFFLAGRSMVWMPIGASLFASNIGSGHFGV 86
 QY 71 TAEAVYVPGYGLAWAQAPIGYS-----LSLILGGLFFAKPMRSKGYVTMLDPFQOIY-GK 124
 Db 87 LA-----GTGAASGIATGGFEWNNALVVLVGLWLFV--GYVTMPYLOKRFQGGK 139
 QY 125 RMGGLLFFPALMGEMFW--AAAFSALGAT-ISVIIDVDMHISVIISALIAITLTLVGLL 181
 Db 140 RIQIYLSLSLLLYFTKISADIFS--GAIFIQTLGLDIIYVAILLVITGLYITGGL 197
 QY 182 YSAVTDVQLFCIFVGLWISVPFALSHPAVDIGTAVHAKY-----Q 225
 Db 198 AAVYITDILQAIMVGVSVLTGFAPFHEVG-----GYEAFTEKYMRAIPQSISYNTSTIPQ 253
 QY 226 KPWLTGDSSEVYSWLDSFLLMLGGIPW-----QAYFORVLSSSSA 267
 Db 254 KYCTPREDAFHI-----FRDAITGDIPWGLVFGMSILTLWYCTQDQIVQRCLSAKNL 307
 QY 268 TYAQVLSFLAAFGCLVMAIPAILIGAIGASTDWNQTXGLPDP-----KTTTEADMLIP 321
 Db 308 SHVKAGCILCGYLYKVMPELIVMGMVSRILYTDKVCVVPSECEKCYGTRVGCNTNIAPP 367
 QY 322 IVLQILCPYVIFSGGLGAVSAAMSSADSILLSASSMFARNIYQISFRONASDKELVWVM 381
 Db 368 TLVVELMPNGLRGLMLSVMMASLSLTSIFNSASTLFTMDIY-TKIRKASEKELMIAAG 426
 QY 382 RI-TVEVFGASATAMALLTKTYG--LWYLSLDLYI--VIFPQLLCVLFVKGNTNYGAV 436
 Db 427 RLFLMLFLIGISIAWPIVOSAGSQGLFDYIQISTYSLGPPIAAVFAIFLFAFKVRNPGAF 486

Db 365 AY-----PTIVELMPNGLRGLMLSVLASLMSLSIFNSASTLFTMDI 409
QY 364 YOLSFQNASDKIEVWMTITFV-FGASATAMALLTKTVYG--LWYLSDLVI--VIF 418
Db 410 Y-AKVRKRASERELTAGRLFILVIGISIAWVPIVQSAQSGOLFIDYIQISYLSGPPA 468
QY 419 POLLCVLFVKGNTYTGAVAGYVSGFLRI-----TG-----GEPYLY 455
Db 469 AVFLAIFWKRNEFGAEGWGLILGLIGISRMITEFAYGTGSCMEPSNCPITICGVHYLY 528
QY 456 LOPLIFYPGYVDDNGIYNQKPFKTLAMVTSFLNICISYLAKYLFESGTLPPKLDVFD 515
Db 529 FAIILF-----AISFTIVISLTKPI-----PDVHLXR 558
QY 516 AV--VARHSEENMDKILVKNENIKDELALVKKPSMTLSSTFTNKE 561
Db 559 LCWLSRNSKEERID--LDAEENIQ-----EGPKETIEIFQVEKK 598
RESULT 6
33582
A:Glucose cotransporter SGLT1 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 12-Apr-1995 #sequence_revision 12-Apr-1995 #text_change 20-Aug-1999
C:Accession: A53582
R:Lee, W.S.; Kanai, Y.; Wells, R.G.; Hediger, M.A.
J. Biol. Chem. 269, 12032-12039, 1994
A:Title: The high affinity Na(+)/glucose cotransporter. Re-evaluation of function and di
A:Reference number: A53582; MUID:94216314; PMID:8163506
A:Accession: A53582
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-665 <LEE>
A:Cross-references: GB:U03120; NID:g414571; PID:AAA19015.1; PID:g414572
C:Superfamily: proline carrier protein
Query Match 10.3%; Score 306; DB 2; Length 665;
Best Local Similarity 23.5%; Pred. No. 1e-14;
Matches 155; Conservative 105; Mismatches 242; Indels 158; Gaps 29;
QY 11 IIVFYLLILVIGIAWRTKNSGAESEAIIVGGRIIGLVGGFTMTATWGGYING 70
Db 32 IIVFYVVMVAVGLWAFST-NRGTV---GGFFLAGRSMVWVPIGASLFASNGSHFVG 86
QY 71 TAEAVYVPGYGLAQAQPTGYSL-----LILGLFFAKPMRSK-GYVTMLDPFQOYVK 124
Db 87 LA-----GTGAAGTAGGFEWNLVVFVVLGWFV---PIYIKAGVYTM---PEYLRK 134
QY 125 RMGG-----LLFIPALMGEMFWAAAFISALGATISVIIDVDMHISVILSALIA 172
Db 135 RFGGRRIQIYLSVLSLLYIFTKISADIFSGAIF-----INLALGDIYLAIFILLAIT 188
QY 173 TLYTLVGLYSVAYTDVQLFCIFVGLWISVFPALSHPAVADIGFTAVHAKYOK--PWL- 229
Db 189 ALYTTITGLAAVYITDTLTALMLVGSFLTGTAFREVG---GYAPFMDKYMKAIPTLV 244
QY 230 --GTVD--SSEVYS-WLDSFLLL--MLGIPW-----QAYFORVLSS 264
Db 245 SDGNITVKECTTPRADSFHFRDPTGDMPWPLGIFGLSILALWVWCTQVIVQRCLSA 304
QY 265 SSATYAQVLSFLAAGCIVMAIPAIIIGAIGASTDMNQTAYGLPDP-----KTTTEADM 318
Db 305 KMSHVRAAGCTLCGYLKLPMFLVMPGCMISRIILYTKVACVVPSECKKYCTPGVCTNI 364
QY 319 ILPIVQLVCPVYISFFGLGAVSAVMSADSSILSASSMFARNIYQLSFRONASDKIEV 378
Db 365 AYPLTVVLMPLNGRLMLSVLASLMSLSIFNSASTLFTMDIY-TKIRKASEKELM 423
QY 379 WVMRTITFV-FGASATAMALLTKTVYG--LWYLSDLVI--VIFPQLLCVLFVKGTNTY 433
Db 424 IAGRLFLVLIGISIAWVPIVQSAQSGOLFIDYIQISYLSGPPIAAVFLAIFCKRVNEP 483
QY 434 GAVAGYVSGFLRI-----TG-----GEPYLYLQPLIFYPGYVDDN 470

Db 484 GAFWGLILIGELIGISRMITEFAYGTGSCMEPSNCPKICGVHYLYFAIILF----- 534
QY 471 GIYNQKPFKTLAMVTSFLNICISYLAKYLFESGTLPPKLDVFDVAV--VARHSEENMDK 528
Db 535 -----AISVTVLVISLTKPI-----PDVHLXRLCWSLRNSTEERID- 572
QY 529 TILVKNENIKDELALVKKPSMTLSSTFTNKE-----AFLDVDSPEGSGTGD 577
Db 573 --LDAGEEPVEE---DPKDTIEDAEPQKEKCFKAYDLFCGLDQDQKGPVKMTKEE 626
RESULT 7
A44432
A:amino acid transport protein - pig
N:Alternate names: Na+/amino acid cotransporter, SAAT1
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Aug-1999
C:Accession: A44432
R:Kong, C.T.; Yet, S.F.; Lever, J.E.
J. Biol. Chem. 268, 1509-1512, 1993
A:Title: Cloning and expression of a mammalian Na+/amino acid cotransporter with se
A:Reference number: A44432; MUID:93131881; PMID:8420925
A:Accession: A44432
A:Molecule type: nucleic acid
A:Residues: 1-660 <KON>
A:Cross-references: GB:L02900; NID:g164666; PIDN:AAC37325.1; PID:g164667
A:Experimental source: kidney epithelial cell line LLC-PK1
A:Note: sequence extracted from NCBI backbone (NCBI:122778)
C:Superfamily: proline carrier protein
C:Keywords: amino acid transport; membrane protein
Query Match 10.2%; Score 303.5; DB 2; Length 660;
Best Local Similarity 23.2%; Pred. No. 1.5e-14;
Matches 141; Conservative 103; Mismatches 230; Indels 135; Gaps 26;
QY 11 IIVFYLLILVIGIAWRTKNSGAESEAIIVGGRIIGLVGGFTMTATWGGYING 70
Db 32 IIVFYVVMVAVGLWAFST-NRGTV---GGFFLAGRSMVWVPIGASLFASNGSHFVG 86
QY 71 TAEAVYVPGYGLAQAQPTGYSL-----LILGLFFAKPMRSKGYVTMLDPFQOYI-GKRM 126
Db 87 LAGTGAAGTAAAFW-----NALLLLVLGFMFFIYIKAGVMTPEYLRKRFSGKRL 141
QY 127 GGLFIPAL-----MGEMFWAAAFISALGATISVIIDVDMHISVILSALIAITLV 179
Db 142 QIYLSLSLFCVALRISSDIFSGAIF-----IKLALGLDLYLAIFSLAITAITITG 195
QY 180 GLYSVAYTDVQLFCIFVGLWISVFPALSHPAVADIGFTAVHAKYOK--PWLGTVD----- 233
Db 196 GLASVITDTLTQITIMLIGSFLMGFAF---VEVGGVESFTEKYMNAIPTIVEGDNLT 251
QY 234 SSEVYS-WLDSFLLL--MLGIPW-----QAYFORVLSSSATYAO 271
Db 252 SPKCYTPQGSFHFIRDAVTDGIDPWPFGMIFGVVVAWVWCTQVIVQRCLSGKDMSHVK 311
QY 272 VLSFLAAGCIVMAIPAIIIGAIGASTDMNQTAYGLPDPKT-----TEE--ADMILPIV 325
Db 312 AACIMCYLKLPMFLVMPGCMISRIILYTKVACVVPSECKKYCTPGVCSNAYPLVW 371
QY 326 YLCPVYISFFGLGAVSAVMSADSSILSASSMFARNIYQLSFRONASDKIEVWMTITV 385
Db 372 ELMPGSLRGLMLSVLASLMSLSIFNSASTLFTMDIY-TKIRKQASEKELLIAAGRLFI 430
QY 386 FVEGASATAMALLTKTVYG---LWYLSDLVI--VIFPQLLCVLFVKGTNTYGA-----V 436
Db 431 ILLIVISVWVPLVQVAQNGOLFHYIESISSYLGPPIAAVFLAIFCKRVNEQGAFWGLI 490
QY 437 AGVVSGL-----FLRITG-----GEPYLYLQPLIFYPGYVDDNGIYNQK 477
Db 491 IGFVMDGLIRMAEFVYGTGSCLAASCPQICGVHYLYFAIILF----- 535
QY 478 PFKTLAMVTSFLNICISYLAK-----YLFE-----SGLTPPKLDVFDVAVRH----- 521

Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: B83988
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-507 <STO>
A:Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BA06425.1; GSPDB:
A:Experimental source: strain C-125
C:Genetics:
A:Gene: opuE
C:Superfamily: proline carrier protein

Query Match 10.1%; Score 299.5; DB 2; Length 507;
Best Local Similarity 26.2%; Pred. No. 2.2e-14;
Matches 141; Conservative 84; Mismatches 220; Indels 93; Gaps 27;

QY 5 VEGE-IAIVFYLL-ILLVGWAARTKNSGAERSEAIIVGGRDIGLLVGGFTMTATW 62
DB 4 VEPLAVAILIAVLALLIGLSS-KKSVGMTD-----FFIAGRNLNKWTVALSAVSSG 57
QY 63 VGGYINGTAEAVYVPGYGLAWAQAPIGYSLILGLGFFAKPMRSKY-----VTMLD 116
DB 58 RSANLVLTGTATGLDANWAVA--GIITVEVF--LEFYVARRFRAISEQTGSITPD 113
QY 117 PFQIYVGR---MGGLLFIPALMGEMFWAAIFSAI---GATISVIIDVDMHISVISA 169
DB 114 ILETRENDKTHILRGGSAFI--IM--FFMIAVVASOLVAGGAFATSMGVSSSTGMWVA 169
QY 170 LIATLVTLVGLYSVAYTDVQVLCIFVGLWISVPPALSHPAVADIG-----FTAVHA 222
DB 170 VILLAYTLMGGHVAVKTDVQAGFMFVSLVIL-----PVVAILGLGDFDPLLOVMHT 222
QY 223 KYQKPMWGTVDSEVYSWLDLMLG-GIPWQAY-FQVLSSTSATYAQVLSFIAFG 280
DB 223 EG-----GGTSPFAFGFAGVIGLLGFGSPGNPHILVRYMSLVNKKVEMQAALISSVW 277
QY 281 CLVMAIPAILIGAISTDMNOTAYGLDPDKTTEAD--MILPIVQLVCLPVYISFGL 337
DB 278 NVLMGNGAVMIGLAG-----RAY-FPDVSLLPNGDEQVFLMLGSEILHPLFFGL-L 328
QY 338 GAVSAVMSADSSILSASSMEARNIYQLSFRON--ASDKEIVMVMRITVVFEGASATAM 395
DB 329 VAVLAAMSAQSLLVGSFAFVDIYQKMRNRKLSQKLVRLSLTTVWFVGLSLIL 388
QY 396 ALLTKTVYGLWYSSDLVYIVF-----PQLCVLFVKGTNTYTGAVGVSGFL 445
DB 389 A-PTAQEEFVW-----NVLFAGGIGAGCFGAPALLSFYKMGVTRQGVLMGMIAGLLT 439
QY 446 RITGGEPLVLOPLIFPGYVDDNGIYNOKFPEKTLAMVTSFLT-----NICISYLAK 499
DB 440 VI-----LVKQOQOWTY-AFLPDVKELNNTYFVGITVEAVPGFIVATTITVVISLETK 491

RESULT 10
A42251
nucleoside transport protein - rabbit
N:Alternate names: Na+/nucleoside cotransporter, SNST1
C:Species: *Oryctolagus cuniculus* (domestic rabbit)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Aug-1999
C:Accession: A42251
R:Pajor, A.M.; Wright, E.M.
J. Biol. Chem. 267, 3557-3560, 1992
A:Title: Cloning and functional expression of a mammalian Na+/nucleoside cotranspor
A:Reference number: A42251; MUID:92156077; PMID:1740408
A:Accession: A42251
A:Molecule type: mRNA
A:Residues: 1-672 <PAJ>
A:Cross-references: GB:M84020; NID:g165550; PIDN:AAA31421.1; PID:g165551
A:Note: sequence extracted from NCBI backbone (NCBI:82253, NCBI:82256)
C:Superfamily: proline carrier protein
C:Keywords: membrane protein; nucleoside transport

Db 536 -----VSILVVLAISLTTPDPVHLRLCWALRNSTEERIDL-DAEKRHEEAHDG 586
QY 522 -SEENMDKT 529
Db 587 VDEDNPET 595

RESULT 8
E83468
probable sodium/solute symporter PA1418 [imported] - *Pseudomonas aeruginosa* (strain PA01
C:Species: *Pseudomonas aeruginosa*
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83468
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: E83468
A:Status: preliminary
A:Molecule type: DNA
A:Cross-references: GB:AE004571; GB:AE004091; NID:g9947360; PIDN:AAG04807.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1418

Query Match 10.1%; Score 301; DB 2; Length 463;
Best Local Similarity 25.1%; Pred. No. 1.5e-14;
Matches 115; Conservative 86; Mismatches 211; Indels 46; Gaps 15;

QY 9 IAIIVFYLLLVGI-----WAAWRTKNSGAERSEAIIVGGRDIGLLVGGF---TMTAT 61
DB 1 MALDIFVLLIYAAGMTALGWGNR-----RAKTRDD-YLVAGRNLG---PGFVLGTMAAT 51
QY 62 WVGGYINGTAEAVYVPGYGLAWAQAPIGYSLILGLGFFAKPMRSKYVYVMDLDFQOI 121
DB 52 VLGGASTIGTVRLGYVHGIGSEFWLCAIG--IGIVGLSLFLAKPLKLIYTVQVLERR 109
QY 122 YGRMGGLFIPALMGEMFWAAIFSAIGATISVIIDVDMHISVIALIATLYTLVGL 181
DB 110 YNPAARHASALIMLVALMIGATSTTAIGTVMQVILGLFPWVSILIGGGVVLYSTIGM 169
QY 182 YSVAYTDVQVLCIFVGL-WISVPFALSHPAVADIGFTAVHAKYQKPLWLTVDSEVYSW 240
DB 170 WSLTLTDIVQFLIMTVGLVFLMLPSINDAG---GWDALVAKLPASYF---DFTAI-GW 221
QY 241 ---LDSELLMLGIPQWYFORVLSSTSATYAQVLSFLAAGGLVMAIPAILIGAIGAS 297
DB 222 DTIVTVFLIYFFGIFIGODIQRVFTARSTVAKVAGSAGIYCVLYGMAGALIGMAKV 281
QY 298 TDMNOTAYGLPDKPTTEADMLPIVQLVCLPVYISFGLGAVSAVMSADSSILSASS 357
DB 282 L-----LPD-----LENNVNAFASVVEHSLPNCIGRLVIAAAALAMSTASAGLLAAST 330
QY 358 MFARNIY-QLSFRONASKEIVWMRTVVFVFGASATAMALLTKTVGLWYSSDLVIV 416
DB 331 TVTQDLPLRLRRGQSDNGVDHENRIATLLGLVGLVIALVWSDVISALTVAYNLLVGG 390
QY 417 IFPOLLCVLFVKGTNTYGAFA-----GVVSGFLRITGG 450
DB 391 MLIPILGAIYWKRAATTAGAITSMVLGFLTVLVFMKDG 428

RESULT 9
B83988
proline transporter opuE [imported] - *Bacillus halodurans* (strain C-125)
C:Species: *Bacillus halodurans*
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: B83988
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Query Match 10.0%; Score 298; DB 2; Length 672;
 Best Local Similarity 25.0%; Pred. No. 3.9e-14;
 Matches 153; Conservative 89; Mismatches 232; Indels 138; Gaps 25;

QY 9 IAIIVFVLLILLVCIWAARTKNSGSAERSEALIVGGDGLVGGFTMTATWVGQY 67
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 26 IAVIAAFLVIGVGLNSMCRN-NGTV-----GGYFLAGSMVWMPVGSALFASNIGSGH 80

QY 68 INGTAEAIVVPGYGLAWAQAIPIGYSLS-----LIIGLFFFAKPMRSKGVVMTLDPFQOY 123
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 81 FVGLA-----GTGAANGLAVAGFEWNAFVVLGLLWLFAPVLTAGVITM-----POYL 130

QY 124 KRMGV-----LLFIPALMGEMFWAAAF-----SALGATISVIIDVDMHISVITSA 169
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 131 KRGGHRIRLYLSVLSFLIFTKISVDMFSGAVFIQOALGWN-----YASVIAL 182

QY 170 LIATLYTLVGLSVAYTDVVQLFCIFVGLWISVPEALSHPAVDIGTAVIAKY----- 224
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 183 GITWVTVTGGLAALMTDTVQTFVIITAGAFILTYAFHEVG-----GYSGLFDKYMGAMT 238

QY 225 -----OKPWLGTVDSDSEVSWLDSFLL-----MLGGIPW-----QAYF 258
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 239 SLTVSEDPVAGNISSSCYRPRPSYHLRDPVTGDLFPALLLGLTIVSGWYWCSDQVIV 298

QY 259 QRYLSSSSATYQVLSFLAFAFGCLVMAIPAILIGAGSTDNQNTAYGLDPKTI-----TE 314
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 299 QRLCAGRLNTHIKAGCILCYLKLTPMLVMFGMISRIILYDEVACVPECKRVCGTE 358

QY 315 E--ADMILPVLQYLCPVISFGLGAVSAAMVSSADSSILSSASSMFARNIYQLSFRONA 372
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 359 VGCNITAYPRLVVVKLPNGLRLGLMLAVMLAALMSSLASIFNSSTLTFTMDIYL--RPA 416

QY 373 SKKEIVWMRITVVFAGSATAMALLTKTVYG---LWYSSDLVYI--TFPOLLCLVLFV 427
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 417 GEGELLVGLWVFIIVAVSWLPPVQAAGQLFDYIQSVSSYLAAPPVSAVFVVALFV 476

QY 428 KFTNYGAVGVSGFLRITGEPYLYLQPLIFYPGYPPDDNGIYNQKPFKTLAMV-- 485
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 477 PRVNEKGFAGWLGILGLMLARLP-----EFSGTGSQVRP 513

QY 486 ---TSFLTNICISYLAKEYLFE-SG-----TLP-PKLDVFPDVAVA-RHSEENMDKTI 530
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 514 SACPAFLCRVHLYFAIVLFFCSGLLIITVSLCTAPIPRKHHLRLVFLSHSKE----- 567

QY 531 LVKNENIKLDEL 542
 : : : : :
 Db 568 --EREDLDADDEL 577

RESULT 11
 5596337
 glucose transport protein SGLT1, intestinal - sheep
 N:Alternate names: Na+/glucose cotransporter SGLT1
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 20-Aug-1999
 C:Accession: S59637; S48858
 R:Tarpey, P.S.; Wood, I.S.; Shirazi-Beechey, S.P.; Beechey, R.B.
 Biochem. J. 312, 293-300, 1995
 A:Title: Amino acid sequence and the cellular location of the Na(+)-dependent D-glucose
 A:Reference number: S59637; MUID:96077158; PMID:7492327
 A:Accession: S59637
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-664 <R>
 A:Cross-references: EMBL:X82411; NID:g861072; PIDN:CAA57809.1; PID:g861073
 A:Experimental source: tissue type jejunal mucosa
 R:Wood, I.
 submitted to the EMBL Data Library, October 1994
 A:Reference number: S48858
 A:Accession: S48858
 A:Molecule type: mRNA
 A:Residues: 1-233, 'R', 235-432, 'V', 434-466, 'MR', 469-664 <WOO>
 A:Cross-references: EMBL:X82411

C:Superfamily: proline carrier protein

Query Match 9.9%; Score 294; DB 2; Length 664;
 Best Local Similarity 23.9%; Pred. No. 7.6e-14;
 Matches 127; Conservative 93; Mismatches 202; Indels 110; Gaps 23;

QY 11 IIVFYLLILLVGIWAARTKNSGSAERSEALIVGGDGLVGGFTMTATWVGQY--- 67
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 32 IIVFYVWAVGLWAMEST-NGTV-----GGYFLAGSMVWMPVGSALFASNIGSGH 86

QY 68 INGTAEAIVVPGYGLAWAQAIPIGYSLSLIIGLFFFAKPMRSK-CYVMTLDPFQOYIKRM 126
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 87 LAGTGAAGIATGGFEW-----ALILVLLGLWVFW---PIYKAGVVTM-----PEYLKRF 136

QY 127 GG-----LLFIPALMGEMFWAAAFISALGATISVIIDVDMHISVITSAIATL 174
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 137 GGRIQVYLSVLSVLYIFTKISADIFSGAIF-----INLGLDLYLAIFILLATL 190

QY 175 YTLVGLSVAYTDVVQLFCIFVGLWISVPEALSHPAVDIGTAVIAKYOKPWLGTVD 234
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 191 YTIITGLAAVYITDTLTQTVIMLIGSFILGTAFHEVG-----GYSAFVTKYMA-IPVTS 245

QY 235 -----SEVYS-WLDSFLL-----MLGGIPW-----QAYFORVLSS 265
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 246 YGNTVTKKCYTPRADSPHIFRDLPLKGLPWCLIFGLFTIISLWYCTDOVIVQRCUSAK 305

QY 266 SATYAQVLSFLAFAFGCLVMAIPAILIGAGSTDNQNTAYGLDPKTEE-----AD 317
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 306 NWSHVKACIMCGYMKLLPMLFWMVPMGMSIRILFTEKVACTV--PSECEKYCGTKVGCTN 363

QY 318 MLPIVLQYLCPVISFGLGAVSAAMVSSADSSILSSASSMFARNIYQLSFRONASDEKI 377
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 364 IAYPLTVVELMPNGLRLGLMLSVMLASLSLTSIENSASTLTMDIY-TKIRKASEKEL 422

QY 378 VVWRITVYFV-FGASATAMALLTKTVYG---LWYSSDLVYI--VIFPOLLCLVLFVKGNT 432
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 423 MIAGRLFMVLVIGSIWVPIVQSAQGLFDYIQSITSYLGPPIAAVFLLAIFCKRVNE 482

QY 433 YGAVAGVSGFLRIT-----TG-----GEPYLYLQPLIF 461
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 483 PGAFWGLIIGLIGVSRMITFAFGTSCMEPSNCPPTICGVHYLYFAILF 534

RESULT 12
 A56765
 sodium-glucose cotransporter homolog - human
 C:Species: Homo sapiens (man)
 C:Date: 08-Sep-1995 #sequence_revision 08-Sep-1995 #text_change 20-Aug-1999
 C:Accession: A56765; I51890
 R:Wellis, R.G.; Pajor, A.M.; Kanai, Y.; Turk, E.; Wright, E.M.; Hediger, M.A.
 Am. J. Physiol. 263, F459-F465, 1992
 A:Title: Cloning of a human kidney cDNA with similarity to the sodium-glucose cotran
 A:Reference number: A56765; MUID:93035768; PMID:1415574
 A:Accession: A56765
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-672 <WEL>
 A:Cross-references: GB:M95549; NID:g338052; PIDN:AAA36608.1; PID:g338053
 A:Experimental source: kidney cortex
 C:Superfamily: proline carrier protein
 C:Keywords: transmembrane protein

Query Match 9.8%; Score 292; DB 2; Length 672;
 Best Local Similarity 24.1%; Pred. No. 1.1e-13;
 Matches 147; Conservative 91; Mismatches 237; Indels 136; Gaps 22;

QY 8 LIAIIVFYLLILLVGIWAARTKNSGSAERSEALIVGGDGLVGGFTMTATWVGQY 67
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 26 ILVIAAYFLVIGVGLNSMCRN-NGTV-----GGYFLAGSMVWMPVGSALFASNIGSGH 80

QY 68 INGTAEAIVVPGYGLAWAQAIPIGYSLS-----LITGLLFFAKPMRSKGVVMTLDPFQOY 123
 ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
 Db 81 FVGLA-----GTGAASGLAVAGFEWNAFVVLGLLWLFAPVLTAGVITM-----POYL 130

124 KRMGG-----LLFIPALMGEMFWAAAI--SALGATISVIIDVDMHISVIISA 169
 131 KRFGRRIRLYLSVLSFLYIFTKISADIFSGAIF-----INLALGLDLYLAIFLILAITAL 190
 170 LIATLYTLVGLKSVATDVQVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPMWLGTVDS 234
 183 GITMIYTVGGLAAMTDTVQTFVILGGACILMGYAFHEVG-----GYSLFPDKYLGAAT 238
 225 -----QKPMWLGTVDSSEVYSWLDSEFL--MLGGIPW-----QAYF 258
 239 SUTVSEDPVAGNISFCYRPRDSYHLLRHPVTGDLFPALLGLTIVSGWYWCSDQVIV 298
 259 QRVLSSTATYQVLSFLAAGCLVMAIPAILIGAIGASTDWNQATAYGLPDPKTE-----AD 317
 299 QRLAGKSLTHIKAGCILGKYLKTFMFLMYPMGIMSRILYFTEKRVACTV--PSECEKCYGTVKGCN 363
 315 E--ADMILPVLQVLCPIVIFSGFLGAVSAVMSADSSILSASSFARNIYQLSFRQNA 372
 359 VGCNIAYPRLVVKLPNCLRGCLMLAVLMAALMSLASFNSSTLTFTMDIY--TRLPRA 417
 373 SKKEIVVMRI--TVFVGASATAMALLTKTVYGLWLSDDLVIYIRPOLLCV-----LFV 427
 418 GDRELLVGLWVFIIVVSVAWLPVQAQGGQLFDYIQAVSSVLAIPVSAVFLAVF 477
 428 KGTNTYGVAGYVSGFLRAITGEPVLYLQPLIFEPGYPPDNGTYNQKFPFKTLAMY-- 485
 478 PRVNEQAGFWLGGGLGMLARLIP-----EFSGSGCVQP 514
 486 ---TSFTNLCISYLAKYLFE--SGTLPPKLVDFDAV-----ARSEENMDKTI 530
 515 SACPAFLCGVHYLYFAVLFPCSGLLTLVSLCTAPIPRKHLRLVFSLRHSE----- 568
 531 LVKNENIKLDE 541
 569 --EREDLDADE 577

RESULT 13
 S59638
 N:Alternate names: Na+/glucose cotransporter SGLT1
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 07-May-1999
 C:Accession: S59638; S48857
 B:Tarpey, P. S.; Wood, I. S.; Shirazi-Beechey, S. P.; Beechey, R. B.
 Biochem. J. 312, 293-300, 1995
 A:Title: Amino acid sequence and the cellular location of the Na(+)-dependent D-glucose
 A:Reference number: S59637; MUID:96077159; PMID:7492327
 A:Accession: S59638
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-664 <TAR>
 A:Cross-references: EMBL:X82410
 A:Experimental source: clone SGLT1B; tissue type parotid gland
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994
 C:Superfamily: proline carrier protein
 C:Keywords: transmembrane protein

Query Match 9.7%; Score 288; DB 2; Length 664;
 Best Local Similarity 23.7%; Pred. No. 2.1e-13;
 Matches 126; Conservative 93; Mismatches 203; Indels 110; Gaps 23;

11 IVFVLLILVLTWAAWRKNSGSAERSEAIIVGGRDGLLVGGFTMTATWVGGY--- 67
 32 IVIVFVVMVAVGLWHMFST--NRGT-----GGFFLAGRSVMWNPVIGASLFSNIGSGHFGV 86
 68 INGTAEAVVPGYGLAWAQAAPICYSLSLILGGLFFAKPMRSK--GYVTMLDPPQQIYGRM 126
 87 LAGTGAAGATGGFEWN-----ALLVLLGWVFEV--PIYKAGVVTM-----PEYLRKRF 136
 127 GG-----LLFIPALMGEMFWAAAI--SALGATISVIIDVDMHISVIISAIALTL 174

137 GGQIQIVLSVLSLVLYIFTKISADIFSGAIF-----INLALGLDLYLAIFLILAITAL 190
 175 YTLVGGLYSVAYTDVQVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPMWLGTVDS 234
 191 YTTIGGLAAVYTTDTLOTQVIMLLGSLFILTGFAPFHEVG-----GYSAFTYKYMNA--IPVTS 245
 235 -----SEVYS-WLDSFLL--MLGGIPW-----QAYFORVLSSS 265
 246 YGNTTVKKECTYPRADSFIHFRDPLKGLDPLWPGILFGLTIISLWYWCMTDQVIVQRCUSAK 305
 266 SATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQATAYGLPDPKTE-----AD 317
 306 NMHVKAGCIMGYMKLLPMFLMMPGIMSRILYFTEKRVACTV--PSECEKCYGTVKGCN 363
 318 MILPVLQVLCPIVIFSGFLGAVSAVMSADSSILSASSFARNIYQLSFRQNASKEI 377
 364 IAYPTLVLEMPNGLRGLMVLASLMSLSLTSIFNSASTLTFTMDIY--TKIRKASEKEL 422
 378 VVMYRITVEV--FGASATAMALLTKTVYGLWLSDDLVIY--VIFPOLLCVLFVKGNT 432
 423 MIAGRLFMVLVIGVSTAWPVIQSAQSGQLFYIOISITSLYGLPPTANVFLAIFCKRVNE 482
 433 YGAVAGYVSGFLFRI-----TG-----GEPYLYLOPLIF 461
 483 PGAFWGLIIGLIGVSRMITEFAYGTCGMEPSNCPCTIICGVHYLYFAILF 534

RESULT 14
 H71097
 hypothetical protein PH1044 - Pyrococcus horikoshii
 C:Species: Pyrococcus horikoshii
 C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
 C:Accession: H71097
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hino, Y.; Yamamoto, S.; Ogura, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Ogura, Y.; Onofu, Y.; Funahashi, T.
 DNA Res. 5, 55-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
 A:Reference number: A71000; MUID:98344137; PMID:9679194
 A:Accession: H71097
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-491 <KAW>
 A:Cross-references: GB:AP000004; NID:g3236131; PIDN:BAA30142.1; PID:g3257459
 A:Experimental source: strain OT3
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
 C:Genetics:
 C:Gene: PH1044
 C:Superfamily: proline carrier protein

Query Match 9.6%; Score 286; DB 2; Length 491;
 Best Local Similarity 22.9%; Pred. No. 2.1e-13;
 Matches 125; Conservative 83; Mismatches 197; Indels 142; Gaps 20;

8 LIAIVFVLLILVLTWAAWRKNSGSAERSEAIIVGGRDGLLVGGFTMTATWVGGY 67
 17 LITVGLWVFLSVLVGYMAGIKRFT-----LEGYLSGRTGLIFLYVLMAGEIYSAYA 70
 68 INGTAEAVVPGYGLAWAQAAPICYSLSLILGGLFFA-----KPMRSKGYVTMLDPPQQIY 123
 71 FLGTHGWAYSYGMPINVA---ICYGALAYSFGYFYARYVMKAGKFCGCVTOADYFOVRN 127
 124 KRMGGLLFIPLMGEMF---WAAAI--SALGATISV---IIDVDMHISVIISAIALTL 178
 128 SK--ALAVLVAIGIIFNIPYLOQLGGLYIVHVSGLSITPKAGIVGIMMIIMVYVT 185
 179 GGLYSVAYTDVQVQLFCIFVGLWISVPFALSHPAVADIGFTAVHAKYQKPMWLGTVDSSEVY 238
 186 SGLRGISWNILQATLMFIYAVV--VLTIPKPGGGIGELFKTLAOTKP----- 233
 239 SWLDSFLLILGGLGIPWQAYFORVLSSSNTYAQVLSFLAAGCLV--MAIPAILIGAIGA 296
 234 ---DHILHPPILGISW-----YVSTL--ILSLGLGFFWYQPLFPI----- 268

Db 181 IVGLLAITAVYTVAGGLAAVYTDALQTLMLGALTLMGY--SFAAVG--GMEGLKEY 236
QY 225 -----QKPMGLGTVDSEVYS-WLDSFLLLMLGGI 252
Db 237 FLALASNRSENSSCGLPREDAFHIFRDLPLSDLPWPGVLFMGSIPLWY----- 285
QY 253 PW---QAYFORVLSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDNNOTAYGLPD 309
Db 286 -WCTDQVIVQRTLAANKLSHAKGALMAAYLKVLPLFTMPVPGMVSRLFPDQVA--CAD 342
QY 310 PKTEE-----ADMILPIVLOQLCPVYISFFGLGAVSAVMSADSSILSASSMPEAR 361
Db 343 PEICQKICSNPSCGSDIAYPKLVLELLPTGLRGLMMAVMAALMSSLTSIENSASTITFM 402
QY 362 NIYOLSPRONADKEIWMVRITVVFEGASATAMALLTKTVYGLW-----Y 407
Db 403 DLWN-HLRPRASEKELMIVGRFV-----LLVLVSILMIPVQASQGGOLFII 450
QY 408 LSSDLVYI---VIFPQLLCVLFVKGNTYGVAGVYVSGFLFRITG-GEPLYLQPLIF 461
Db 451 IQSISSTYLOPPVAVF---IMGCEWKRTNEKAGFWGLISGLLLGLVRLVLDIYVQPRC- 506
QY 462 YPGYPPDNGIYNQKFFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV----- 513
Db 507 ---DQDERPVLVKSIIHYLFESMILSTVTLITVTSVWF-----TEPPSKEMVSHLTWFT 558
QY 514 -FDAVVARHSEENMDKTLVKNENIKLD-----ELALVKPRQSMTSLTSFTNKEA 562
Db 559 RHDVPVQKEQAPPAAPLSLTLSONGMPASSSSSVQFEMVQENTSKTHSCDMPKOS 615

RESULT 4

US-10-162-012-27

; Sequence 27, Application US/10162012
; Publication No. US20030051660A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Silos-Santiago, Inmaculada
; TITLE OF INVENTION: NOVEL HUMAN ION CHANNEL AND TRANSPORTER FAMILY MEMBERS
; FILE REFERENCE: 10448-190001
; CURRENT APPLICATION NUMBER: US/10162,012
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 60/209,845
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: US 09/875,321
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US01/18340
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/209,257
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,423
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18398
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/209,238
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,363
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18247
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/227,068
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 09/928,530
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/25475
; PRIOR FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: US 60/226,770
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/934,421
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26096
; PRIOR FILING DATE: 2001-08-21

; PRIOR APPLICATION NUMBER: US 60/279,281
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 10/109,029
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: PCT/US02/09728
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/290,288
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US (not assigned)
; PRIOR FILING DATE: 2002-05-13
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 675
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-162-012-27

Query Match 10.0%; Score 298.5; DB 9; Length 675;
Best Local Similarity 22.7%; Pred. No. 1.7e-17;
Matches 149; Conservative 115; Mismatches 238; Indels 155; Gaps 29;

QY 2 AFHVEGL---IAIVFY-LLILVGIWAAMRTKNSAERSEAIIVGGRDIGLLVCGF 56
Db 18 AFPQKGLEPGDIAVLVLYFLVLAAGLWSTVKTR---DVTKGYFLAEGNMVMPVGA- 72
QY 57 TMTATVGGGYINGTAEAVVPGVGLAWAQAIPICYSLS---LILGLFFAKPMRSKY 111
Db 73 SLFASNVGSGHFHGLA-----GSGAATGVSAYELNGLFSVLMWIFL--PIYIAGO 124
QY 112 VTMLDPFQOYIKRMGGLLIPALMGEMFWAAAFISALGATI---SVIID---VDMHIS 164
Db 125 VTTMPEYLR---KREGGIR-IPILAVLYFIYIFTKISVDMYAGAIQQSSHDLVLA 180
QY 165 VIISALITATLYLVGGLYSVAYTDVQVLCIFVGLWISVPALSHPAVDIGFTAVHAKY 224
Db 181 IVGLLAITAVYTVAGGLAAVYTDALQTLMLGALTLMGY--SFAAVG--GMEGLKEY 236
QY 225 -----QKPMGLGTVDSEVYS-WLDSFLLLMLGGI 252
Db 237 FLALASNRSENSSCGLPREDAFHIFRDLPLSDLPWPGVLFMGSIPLWY----- 285
QY 253 PW---QAYFORVLSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDNNOTAYGLPD 309
Db 286 -WCTDQVIVQRTLAANKLSHAKGALMAAYLKVLPLFTMPVPGMVSRLFPDQVA--CAD 342
QY 310 PKTEE-----ADMILPIVLOQLCPVYISFFGLGAVSAVMSADSSILSASSMPEAR 361
Db 343 PEICQKICSNPSCGSDIAYPKLVLELLPTGLRGLMMAVMAALMSSLTSIENSASTITFM 402
QY 362 NIYOLSPRONADKEIWMVRITVVFEGASATAMALLTKTVYGLW-----Y 407
Db 403 DLWN-HLRPRASEKELMIVGRFV-----LLVLVSILMIPVQASQGGOLFII 450
QY 408 LSSDLVYI---VIFPQLLCVLFVKGNTYGVAGVYVSGFLFRITG-GEPLYLQPLIF 461
Db 451 IQSISSTYLOPPVAVF---IMGCEWKRTNEKAGFWGLISGLLLGLVRLVLDIYVQPRC- 506
QY 462 YPGYPPDNGIYNQKFFKTLAMVTSFLTNICISYLAKYLFESGTLPPKLDV----- 513
Db 507 ---DQDERPVLVKSIIHYLFESMILSTVTLITVTSVWF-----TEPPSKEMVSHLTWFT 558
QY 514 -FDAVVARHSEENMDKTLVKNENIKLD-----ELALVKPRQSMTSLTSFTNKEA 562
Db 559 RHDVPVQKEQAPPAAPLSLTLSONGMPASSSSSVQFEMVQENTSKTHSCDMPKOS 615

RESULT 5

US-10-119-988-8
; Sequence 8, Application US/10119988
; Publication No. US20030054453A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.


```

; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: US 09/875,321
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US01/18340
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/209,257
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,423
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18398
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/209,238
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,363
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18247
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/227,068
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 09/928,530
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/25475
; PRIOR FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: US 60/226,770
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/934,421
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26096
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/279,281
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 10/109,029
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: PCT/US02/09728
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/290,288
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US (not assigned)
; PRIOR FILING DATE: 2002-05-13
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-162-012-30

Query Match      9.8%; Score 292; DB 9; Length 672;
Best Local Similarity 24.1%; Pred. No. 6.3e-17;
Matches 147; Conservative 91; Mismatches 237; Indels 136; Gaps 22;

Qy 8 LIAIVFYLLILVGIWAARTKNSGSAERSEAIIVGGRIDIGLLVGGTMTATWVG 67
Db 26 ILVIAAYFLVIGVGLSMCRT-NGRTV---GGYFLAGRSMWMPVGLASFASNGSH 80

Qy 68 INGTAEVVPGYGLAWAQAIPIGYSLSL---LILGGLFFAKPMRSKGYVTMLDPFQIYG 123
Db 81 FVGLA-----GTGAASGLAVAGFENALFVLLGLWLPAPVLTAGVITM-----PQYLR 130

Qy 124 KRMGG-----LLFIPALMGEMFAAIF--SALGATISVIIDVDMHISVIISA 169
Db 131 KRFGRRIRLYSLVSLFLYFTKISVDMSGAVFIQOALGWN---YASVIAL 182

Qy 170 LIATLYLVGLISVAYTDVQLFCIFVGLMISVPFALSHPAVADIGFTAVHAKY----- 224
Db 183 GITMIYVTGGLAALMYDTVOTFVILGGACILMGIAFHEVG-----GYSGLFDKYLGAAT 238

Qy 225 -----QKPLWGLTVDSSEYISWLDSPFL--MLGIPW-----QAYF 258
Db 239 SLTVSEDAVGNISFCYRPDPDSYHLLRHFPVTGDLPPWALLGLLTIIVSGWYWCSDQIV 298

Qy 259 QRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQATAYGLPDKT-----TE 314
Db 259 QRVLSSSSATYAQVLSFLAAGCLVMAIPAILIGAIGASTDWNQATAYGLPDKT-----TE 314

; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: US 09/875,321
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US01/18340
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/209,257
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,423
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18398
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/209,238
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: US 09/875,363
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: PCT/US01/18247
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/227,068
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 09/928,530
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/25475
; PRIOR FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: US 60/226,770
; PRIOR FILING DATE: 2000-08-21
; PRIOR APPLICATION NUMBER: US 09/934,421
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: PCT/US01/26096
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/279,281
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 10/109,029
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: PCT/US02/09728
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/290,288
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US (not assigned)
; PRIOR FILING DATE: 2002-05-13
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-162-012-30

; Sequence 2, Application US/10237859
; Publication No. US20030022309A1
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady V. et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001012CON
; CURRENT APPLICATION NUMBER: US/10/237,859
; CURRENT FILING DATE: 2002-09-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Human
; US-10-237-859-2

Query Match      9.3%; Score 277.5; DB 9; Length 596;
Best Local Similarity 22.7%; Pred. No. 9.4e-16;
Matches 128; Conservative 91; Mismatches 179; Indels 165; Gaps 22;

Qy 5 VEGLIAIVFYLLILVGIWAARTKNSGSAERSEAIIVGGRIDIGLLVGGTMTATWVG 64
Db 18 VADIIVTVFALNAVAVGWSR-----ASRNTVNGYFLAGRDMTWMPIGASFASSEG 72

Qy 65 GGYI-----NGTAEVVPGY-----GLAWAQAIPIGYSLSLILGGLFFAKPMRSKG 110
Db 73 SGLFGLAGSAGAGLAVAGFENATYVLLALAWFVPIYISSEI----- 117

Qy 111 YVTMLDPQOIV-KRGMGLLFIPLMGEMF-----WAAAFSALGATISVIIDVDMH 162
Db 118 -VTLPEYTKRKYGGORIRMYLSVLSLLSVFTKISLDLYAGALF-----VHICLGNFY 170

Qy 163 ISVIFSAIATLYLVGLISVAYTDVQLFCIFVGLMISVPFALSHPAVADI-GFTAVH 221
Db 171 LSTILTLGITALTITAGGLAAVYITDALQTLIMVGVAVI-----LTAKFADIGYGOLE 225

Qy 222 AKYQK-----PWLGTVDSEVYSWLDSPFLMLG 250
Db 226 AAYAQAIPSRITANTTCHLPRDAMHMFDPHTGDLPTWGM-----TFGLTINA 274

Qy 251 GIPW---QAYFORVLSSSSATYAQVLSFLA--GCLVN-----AI 286
Db 275 TWYCTDQIVVORSLSARDLHAKAGSILAYKMLPMGLIIMPMSIRALFPDDVGCYV 334

Qy 287 PAILIGAIGASTDWNQATAYGLPDKTTERADMLPILVQYLCVPIYSFFGLGAVSAVMS 346
Db 287 PAILIGAIGASTDWNQATAYGLPDKTTERADMLPILVQYLCVPIYSFFGLGAVSAVMS 346

```

Db 335 PSECLACAEVGCNSIAY-----PKLVE---LMPIGLRGLM-----IAYMLAALMS 379
QY 347 SADSSILSASSMEARNIYQLSFRQNASDKEIVVMR-ITVVFQASATAMALLTKTYVGL 405
Db 380 SLTISFNSSSTLTMDIWR-RLRPRSGERELLVGRLLVIALIGSVAVIPVLQDSNSGQ 438
QY 406 WYLSDDLVIIVIPQLLCV-----LFVKGNTYGAAGVYSGFLRITG----- 449
Db 439 LFIYMQSVTSSSLAPPVTAFFVLGVFWRRANEQAFWGLIAGLVVGATRLVLEFLNPAPPC 498
QY 450 GEP-----YLYLOPLIF 461
Db 499 GEPDTRPAVLGSIHYLHFAVALF 521

RESULT 9
US-09-740-026A-2
; Sequence 2, Application US/09740026A
; Patent No. US20020081678A1
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady V. et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CLO01012
; CURRENT APPLICATION NUMBER: US/09/740,026A
; PRIOR FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-740-026A-2

Query Match 9.3%; Score 277.5; DB 10; Length 596;
Best Local Similarity 22.7%; Pred. No. 9.4e-16;
Matches 128; Conservative 91; Mismatches 179; Indels 165; Gaps 22;

QY 5 VEGLIAIIVFLLILLVIGWAARFKNSGSAERSEAIIVGGRDGLLVGGFTMTATWYG 64
Db 18 VADIIVTVFALNVAVGINSR-----ASRNTVNGYFLAGRDMTWMPIGASLFASSEG 72

QY 65 GGYI-----NGTAEAVVPGY-----GLAWAQAPIGYSLSLILGLLFPKPMRSKG 110
Db 73 SGLFGLAGSAGGLAVAGFEWNTATVLLALAWVFVPIYISSEI----- 117

QY 111 YVTMLDPFOQIY-KGRMGGLLFIPLMGEMF-----WAAAFSALGATISVIIDVDMH 162
Db 118 -VTLPEYIQRYGQIRMYLSVLSLLSVFTKISLDLYAGALF-----VHICLGWNFY 170

QY 163 ISVIAIALIATLTVLGGLSVATDVVQLFCIFVGLWISVFPFALSHPAVADI-GFTAVH 221
Db 171 LSTILTLGITALYIAGLAAYITDALQTLIMVVGAVI-----LTAKAFDQIGYGQLE 225

QY 222 AKYQK-----PWLGTVDSDSEVYSWLDLSFLLMLG 250
Db 226 AAYAAQAPISRTIANTTCHLPRTDAMHFRDPHTGDLPTWGM-----TFGLTMA 274

QY 251 GIPW---QAYFORVLSSTSSATYQVLSFLAFL-----GCLVM-----AI 286
Db 275 TWVWCTDQIVQVRSLSARDNHAKAGSILASYLKMPLPMGLIIMPMSIRALFPDDVGCW 334

QY 287 PAILGAIGASTDWNQATAYGLDPKTEEADMILPIVLOVLCPVYISFFGIGAVSAVMS 346
Db 335 PSECLACAEVGCNSIAY-----PKLVE---LMPIGLRGLM-----IAYMLAALMS 379

QY 347 SADSSILSASSMEARNIYQLSFRQNASDKEIVVMR-ITVVFQASATAMALLTKTYVGL 405
Db 380 SLTISFNSSSTLTMDIWR-RLRPRSGERELLVGRLLVIALIGSVAVIPVLQDSNSGQ 438

QY 406 WYLSDDLVIIVIPQLLCV-----LFVKGNTYGAAGVYSGFLRITG----- 449

Db 439 LFIYMQSVTSSSLAPPVTAFFVLGVFWRRANEQAFWGLIAGLVVGATRLVLEFLNPAPPC 498
QY 450 GEP-----YLYLOPLIF 461
Db 499 GEPDTRPAVLGSIHYLHFAVALF 521

RESULT 10
US-10-119-988-5
; Sequence 5, Application US/10119988
; Publication No. US20030054453A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Chen, Hong
; TITLE OF INVENTION: Millennium Pharmaceuticals Inc.
; TITLE OF INVENTION: 68723, Sodium/Glucose Cotransporter
; TITLE OF INVENTION: Family Members and Uses Therefor
; FILE REFERENCE: MP101-103plRNM
; CURRENT APPLICATION NUMBER: US/10/119,988
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,764
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-119-988-5

Query Match 9.3%; Score 277.5; DB 9; Length 643;
Best Local Similarity 22.7%; Pred. No. 1e-15;
Matches 128; Conservative 91; Mismatches 179; Indels 165; Gaps 22;

QY 5 VEGLIAIIVFLLILLVIGWAARFKNSGSAERSEAIIVGGRDGLLVGGFTMTATWYG 64
Db 65 VADIIVTVFALNVAVGINSR-----ASRNTVNGYFLAGRDMTWMPIGASLFASSEG 119

QY 65 GGYI-----NGTAEAVVPGY-----GLAWAQAPIGYSLSLILGLLFPKPMRSKG 110
Db 120 SGLFGLAGSAGGLAVAGFEWNTATVLLALAWVFVPIYISSEI----- 164

QY 111 YVTMLDPFOQIY-KGRMGGLLFIPLMGEMF-----WAAAFSALGATISVIIDVDMH 162
Db 165 -VTLPEYIQRYGQIRMYLSVLSLLSVFTKISLDLYAGALF-----VHICLGWNFY 217

QY 163 ISVIAIALIATLTVLGGLSVATDVVQLFCIFVGLWISVFPFALSHPAVADI-GFTAVH 221
Db 218 LSTILTLGITALYIAGLAAYITDALQTLIMVVGAVI-----LTAKAFDQIGYGQLE 272

QY 222 AKYQK-----PWLGTVDSDSEVYSWLDLSFLLMLG 250
Db 273 AAYAAQAPISRTIANTTCHLPRTDAMHFRDPHTGDLPTWGM-----TFGLTMA 321

QY 251 GIPW---QAYFORVLSSTSSATYQVLSFLAFL-----GCLVM-----AI 286
Db 322 TWVWCTDQIVQVRSLSARDNHAKAGSILASYLKMPLPMGLIIMPMSIRALFPDDVGCW 381

QY 287 PAILGAIGASTDWNQATAYGLDPKTEEADMILPIVLOVLCPVYISFFGIGAVSAVMS 346
Db 382 PSECLACAEVGCNSIAY-----PKLVE---LMPIGLRGLM-----IAYMLAALMS 426

QY 347 SADSSILSASSMEARNIYQLSFRQNASDKEIVVMR-ITVVFQASATAMALLTKTYVGL 405
Db 427 SLTISFNSSSTLTMDIWR-RLRPRSGERELLVGRLLVIALIGSVAVIPVLQDSNSGQ 485

QY 406 WYLSDDLVIIVIPQLLCV-----LFVKGNTYGAAGVYSGFLRITG----- 449
Db 486 LFIYMQSVTSSSLAPPVTAFFVLGVFWRRANEQAFWGLIAGLVVGATRLVLEFLNPAPPC 545

QY 450 GEP-----YLYLOPLIF 461
Db 546 GEPDTRPAVLGSIHYLHFAVALF 568

```

; CURRENT APPLICATION NUMBER: US/09/740,026A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
; US-09-740-026A-4

```

```

Query Match          9.3%; Score 277; DB 9; Length 596;
Best Local Similarity 22.5%; Pred. No. 1e-15;
Matches 119; Conservative 92; Mismatches 171; Indels 148; Gaps 20;

```

```

; APPLICANT: MERKULOV, Gennady V. et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CLO01012CON
; CURRENT APPLICATION NUMBER: US/10/237,859
; CURRENT FILING DATE: 2002-09-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
; US-10-237-859-4

```

```

Query Match          9.3%; Score 277; DB 9; Length 596;
Best Local Similarity 22.5%; Pred. No. 1e-15;
Matches 119; Conservative 92; Mismatches 171; Indels 148; Gaps 20;

```

```

QY 5 VEGLIATIVFYLLILVGIWAAR-TKNSGSAERSEALIVGRDI----- 49
DB 18 VTDIVITVYFALNVAVGIWSSCRASRTVS-----GYFLAGRDMTWMPICASLFGSSE 71
QY 50 --GLLVGGFTMTATWGGGYINGTAEVVPVGYGLAWAQAIPGYSLSLILGLLFAKPMR 107
DB 72 GSGLEFG-----LAGSAGAGGLAVA-----GFDWNTATYVLLALAWVFGAIYI----- 113
QY 108 SKGYVTMLDPFOQIY-GKRMGGLLFPALMGEMF-----WAAAFSALGATISVIIDV 159
DB 114 SSEIVTAEYIOKRFGGQIRMYLSVLSLLSVFTKISLDLYAGALF-----VHICLGW 167
QY 160 DMHISVIISALATLYTLVGLYSVAYTDVQVLCIFVGLWISVFPALSHPAVDI-GFT 218
DB 168 NFYLSLTILATLYTITGLVAVIYTDALQTLIMVVGAVI-----LAIKAFHQIDGYG 222
QY 219 AVHAKYOK-----PWLGTVDSSEVYSWLDLSFLLL 247
DB 223 QMEAAVARAIPSRVTANTCHLPRADAMHFRDPYTGDLPTWTCM-----TFGLT 271
QY 248 MLGGIPW--QAYFORVLSSSSATYAQLSFLAAF-----GCLVM----- 284
DB 272 IMATWYCTDQVIVQVRSLSARNLNHAKAGSILASYLKMPLMGLMIMPMSRSLFPDDEVG 331
QY 285 -AIPAILIGAIGASTDWNQAYGLPDKPTTEADMLPIVLYLCPVYISFFGLGAVSAA 343
DB 332 CVVPSECLRACGAEIGCSNIAY-----PKVMELMPVGLRGLMTAVMMPA 376
QY 344 VMSSADSSILSASSMFARNIYQLSFRONASDKIYVWYMRITVFV-FGASATAMALLTKTV 402
DB 377 LMSLSLIFNSSSTLFTMDIWR-RLRPCASERELLVGRVIVVVLIGVSVAVIPVLOGSN 435
QY 403 YGLWYLSDDLVIYVIFPOLLCV-----LFVKGNTNYGAVGVSGLEFLRIT 448
DB 436 GGQLFIYMQSVTSSSLAPPVTAFTLGIFWQRANQGGAFWGLLAGLVAGT 485

```

RESULT 12

```

US-09-740-026A-4
; Sequence 4, Application US/09740026A
; Patent No. US20020081678A1
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady V. et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CLO01012
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CLO01012

```

RESULT 13

```

US-09-738-626-6949
; Sequence 6949, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162

```

;; PRIOR FILING DATE: 2000-04-07
;; PRIOR APPLICATION NUMBER: JP 00/280988
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 7059
;; SOFTWARE: PatentIn ver. 3.0
;; SEQ ID NO 6949
;; LENGTH: 524
;; TYPE: PRT
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6949

Query Match 9.2%; Score 272.5; DB 9; Length 524;
Best Local Similarity 21.7%; Pred. No. 2.1e-15;
Matches 123; Conservative 104; Mismatches 230; Indels 111; Gaps 21;

QY 8 LIAIIVFYLLILVGIWAARTRKNSGAERSEAIIVGGRDIGLLVGGFTMTATWVG 67
DB 8 LIAIIVFYLLILVGIWAARTRKNSGAERSEAIIVGGRDIGLLVGGFTMTATWVG 67
QY 68 INGTAEAVYVPGYGLAWAQAIPGYSLSLIIGLFFAKPMR-----SKGYVTMLDPFOQIY 122
DB 61 LMLPGALFVTGSELM--IAVGLTIGAMAWMWAPRLRSYSISANSITLPSFFENRL 118
QY 123 GRMGGLLPALMGEMFWAAAFSAI---GATISVIIDVDMHISVLSALITATLYLVG 179
DB 119 RDKSRALRIIAALIIIVFFTEYISSGVAGVYWESTFGDYLLGWAIVAGVTVLYTFIG 178
QY 180 GLYSVAYTDVOLFCIFVGLWISV-----FALSHPAVDIGFTAVHAKYQKPMWGLVDSS 235
DB 179 GLFVSYTDVAVOGTINFFSL--IIVPMAYFALANP--MDI--WSFANSNDYGPHTDGINP 234
QY 236 EYVSWLDSF--LLMLGGIPW-----QAYFORVLSSSSATYAQVLSFLAFCGLVM 284
DB 235 TYFSMISGISAIIIGNLGWLGYGPHIVVRFMALRTPAEAKQRRIGISWIIICLG 294
QY 285 AIPAILGAIGASTDNTQAYGLDPKTEE---ADMILPVLQYLCPIVYISFFGLGAVS 341
DB 295 ATFTAILTVFFAQN-----PDANITDTRAYESIFDLARMLFPL--IAGLILTAVL 345
QY 342 AAVMSADSSILSASSMFARNIYQLSFRONASDKETVWVRITVVFVGASATAMALLTKT 401
DB 346 AAINSTMSSQLLVASSLIEDLLKVKKDSLSERTLIMLSRATVILATIAAAMAINP-- 403
QY 402 VYGLWLSLDLYIVIF-----POLLCVLEVKGTNTYCAVAGYVSGFLRTTIGE 451
DB 404 -----SDSILGLVGFAGWAGSAGFPILASLYWKRLNAGASIGMTGAIVSIANG- 455
QY 452 PYLYLOPLI-----FYCPYYPDDNGIYNQKPFKTLAMVTSPLTNICISYLAKYLFESGT 506
DB 456 -----MSPLSDTLYEIIPG-----FALATIVMVV-----VSLITK----- 485
QY 507 LPPKLDVF-----DAVARHSEENMD 527
DB 486 -EPSEILNEFETAKDLAAAVESNEDVD 512

RESULT 14
US-10-119-988-2
; Sequence 2, Application US/10119988
; Publication No. US20030054453A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Chen, Hong
; APPLICANT: Millennium Pharmaceuticals Inc.
; TITLE OF INVENTION: 68723, Sodium/Glucose Cotransporter
; TITLE OF INVENTION: Family Members and Uses Therefor
; FILE REFERENCE: MPI01-103P1RM
; CURRENT APPLICATION NUMBER: US/10/119,988
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 60/282,764
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 2
;; LENGTH: 664
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-119-988-2

Query Match 9.1%; Score 269.5; DB 9; Length 664;
Best Local Similarity 22.1%; Pred. No. 5.3e-15;
Matches 128; Conservative 91; Mismatches 179; Indels 181; Gaps 22;

QY 5 VEGLIAIIVFYLLILVGIWAARTRKNSGAERSEAIIVGGRDIGLLVGGFTMTATWVG 64
DB 65 VADIIVITVFALNVAAGVWSSCR-----ASRNTVNGYFLAGRDMTWMPICASFASSEG 119
QY 65 GGYI-----NGTAEAVYVPGY-----GLAWAQAIPGYSLSLIIGLFFAKPMRSGK 110
DB 120 SGLFICLAGSAGAGGLAVAGFEWNNATVYLLALAWVFVPIVISSEI-----164
QY 111 VYTMLDPOQIY--GRMGGLLPALMGEMF-----WAAATFSALGATISVIIDVDMH 162
DB 165 -VTLPEYIQKRYGGORIRMYLSVLSLSVFTKISLDYAGALF-----VHICLGWNY 217
QY 163 ISVIISALITATLYTLVGGYSVAYTDVOLFCIFVGLWISVFPALSHPAVADI--GFTAVH 221
DB 218 LSTILTIGITALYTIAGGLAAVIYTDALQTLIMVYGAVI-----LTIKAFDQIGGYQGLE 272
QY 222 AKYOK-----PWLGTVDSSEVYSWLDSEFLMLG 250
DB 273 AYAQAIPSRITANTCHLPRTDAMHFRDPTGDLPTGM-----TFGLTMA 321
QY 251 GIPW---QAYFORVLSSSSATYAQVLSFLAAP-----GCLVM-----284
DB 322 TWYCTDQIVVORSLSARDLNAKAGSILASLYKMLPGLIIMPMSIALPFGAIVYEE 381
QY 285 -----AIPAILGAIGASTDNTQAYGLDPKTEEADMLPIVQLYICPV 330
DB 382 RHQVSVSRDVGCVFVSECLRACAGVGCNSIAI---PKLVME---LMPICLRCLM-- 432
QY 331 YISFGLGAVSAAVMSADSSILSASSMFARNIYQLSFRONASDKETVWVR--IYFVVG 389
DB 433 -----IAYMLAALMSSLSITFNSSTLFTMDIWR--RURPSRGERELLVGLRVLVALIG 485
QY 390 ASATAMALLTKYVGLWYLSLDLYIVIFPOLLCV---LFVKGNTNTYCAVAGYVSGFL 445
DB 486 VSWAIPVLQDSNSGQLFIYMQSVTSSILAPPVTAVEVLGVFVWRANEQAGFGLIAGLVV 545
QY 446 RITG-----GEP-----YLYLOPLIF 461
DB 546 GATRLVLEFLNPAPPCGEPDTRPAVLGSIHYLHPFAVALF 584

RESULT 15
US-10-173-123-9
; Sequence 9, Application US/10173123
; Publication No. US20030027301A1
; GENERAL INFORMATION:
; APPLICANT: Hu, Yi
; APPLICANT: Nepomichy, Boris
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Brian
; APPLICANT: Fiddle, Carl Johan
; TITLE OF INVENTION: No. US20030027301A1 Human Transporter Proteins and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0358-USA
; CURRENT APPLICATION NUMBER: US/10/173,123
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/298,241
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 674
; TYPE: PRT

us-10-069-541-6.rapb

Tue Apr 1 13:57:26 2003

ORGANISM: homo sapiens
US-10-173-123-9

Query Match 8.7%; Score 260; DB 9; Length 674;
Best Local Similarity 21.2%; Pred. No. 3.6e-14;
Matches 113; Conservative 87; Mismatches 177; Indels 156; Caps 19;

```
QY 1 MAFHVEGLIALIVFLLILLVGIWAA-----WRTNKGSAEERSE 40
Db 24 VGLHAYDISVVVIVFVFIWISIRASRGITGGYFLAGRSMWMPIGASLMSSNVGS 83
QY 41 AIIIVG-----GRDIGLLVGGFTMTATWGGYINGTAEAVVPGYGLAWAQAPIGYSLSLI 96
Db 84 GLFIGLAGTGAAGGLAVGGFENATWL-----LL 112
QY 97 LGGLFFAKPMRSKGVVTMLDPFOQIYGRMGG-----LLFIPALMGEMFWAAA 144
Db 113 ALGWVFPVPIAAGVVTM---PQYLKKRFGGQRIQVYMSVLSLILYIFTKISTDIFSGA 168
QY 145 IFSALGATISVIIDVDMHISIIISALIALATLYLGGLYSVAKTDVWQ-----LFCIF 196
Db 169 LF-----IOMALGNWLYLSTGILLVTVAVYTIAGGLMAVITYTDALQTVIMVGGALVLMF 222
QY 197 VGL-----W-----ISVPFALSH-----PAVADIGTAVHAKYOK 226
Db 223 LGFQDVGWYPGLEQRYROAIPNVTPTTCHLPRDFAFHMLRDPVSGDI----- 271
QY 227 PWLGTVDSEVYSWLDSEFLMLLGGIPW---QAYFORVLSSSSATYAQVLSFLAAGCLV 283
Db 272 PWPGLI-----FGLTVLATWCWCTDQVIVORSLSAKSLSHAKGGSVLGGYKIL 320
QY 284 MAIPATLIGAIGASTDWNQTAAYCLDPKTEE-----ADMILPVIQYLCPVYISFF 335
Db 321 PMEFFIVMGMISRALFPDEV--GCVPDVCQRICGARVGCNSIAYPKLYMALMPVGLRGL 378
QY 336 GLGAVGAAMVSSADSSILSASSMFARNIYOLSPRONASDKKEIVVMKRI-TVVFEGASATA 394
Db 379 MIAVIMAAALMSSLTISIFNSSSTLFTIDVWQ-RFRKSTEGELMVGVGVFVFLVVISILW 437
QY 395 MALLTKTVYG--LWYLSDDLVIYV--IFPOLLCVLFVKGTNTYGAVAGVYVSGL 443
Db 438 IPIIQSSNSQLEFDYIQAVTSYLAPPITALFLLAIFCKRVTEPCAFWGLVFG 490
```

Search completed: March 31, 2003, 18:48:39
Job time : 595 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 18:34:36 ; Search time 39 Seconds
(without alignments)
437.572 Million cell updates/sec

Title: US-10-069-541-6
Perfect score: 2972
Sequence: 1 MAFHVEGLIAIVFYLLIL.....EAFLDVDSPEGGTNDNLQ 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	308.5	10.4	662	1	US-07-841-651-4
2	298	10.0	672	1	US-07-841-651-2
3	298	10.0	672	1	US-07-841-651-3
4	262.5	8.8	518	4	US-09-134-001C-4744
5	218	7.3	618	4	US-08-595-553A-2
6	206.5	6.9	530	2	US-09-134-001C-4510
7	200	6.7	635	2	US-09-014-969-11
8	138	4.6	482	4	US-09-134-001C-4309
9	123	4.1	496	4	US-09-134-001C-3001
10	119	4.0	499	4	US-09-134-001C-5370
11	116	3.9	832	2	US-08-677-734A-12
12	116	3.9	832	4	US-09-097-053-12
13	113.5	3.8	503	4	US-09-068-195-24
14	110	3.7	599	1	US-08-295-814E-11
15	110	3.7	599	1	US-08-240-783B-4
16	110	3.7	599	3	US-09-084-813-4
17	110	3.7	599	4	US-09-343-361-11
18	110	3.7	599	5	PCT-US92-09662-4
19	109.5	3.7	405	4	US-09-134-001C-4999
20	109	3.7	367	4	US-09-134-001C-5557
21	108	3.6	423	4	US-09-134-001C-3599
22	108	3.6	599	1	US-07-879-617A-10
23	108	3.6	599	1	US-08-753-985-10
24	106	3.6	831	2	US-08-677-734A-11
25	106	3.6	831	4	US-09-097-053-11
26	105.5	3.5	599	1	US-08-301-722A-5
27	104.5	3.5	416	4	US-09-333-208-2

28	104.5	3.5	416	4	US-09-333-254-2
29	104.5	3.5	416	4	US-09-183-270-2
30	104.5	3.5	834	2	US-08-677-734A-9
31	104.5	3.5	834	2	US-08-677-734A-10
32	104.5	3.5	834	4	US-09-097-053-9
33	104.5	3.5	834	4	US-09-097-053-10
34	102.5	3.4	371	2	US-08-928-692-20
35	102.5	3.4	371	4	US-09-339-972-20
36	101.5	3.4	408	4	US-09-134-001C-4130
37	101.5	3.4	1584	4	US-09-251-645-6
38	101	3.4	478	4	US-09-134-001C-4637
39	101	3.4	1296	4	US-09-462-136-9
40	100	3.4	413	4	US-09-134-001C-3702
41	100	3.4	490	4	US-09-134-001C-5116
42	99.5	3.3	437	4	US-09-134-001C-4808
43	99.5	3.3	473	4	US-09-637-118B-2
44	99.5	3.3	522	4	US-09-142-732-2
45	99.5	3.3	635	1	US-07-879-617A-9

ALIGNMENTS

RESULT 1
US-07-841-651-4
; Sequence 4, Application US/07841651
; Patent No. 5410031
; GENERAL INFORMATION:
; APPLICANT: Pajor, Ana M
; TITLE OF INVENTION: Cloning and Functional Expression of a
; TITLE OF INVENTION: Mammalian Na+/Nucleoside Cotransporter: A Member of the
; TITLE OF INVENTION: SGLT Family
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, Ninth Floor
; CITY: Pasadena
; STATE: California
; COUNTRY: USA
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07/841,651
; APPLICATION NUMBER: 19920224
; FILING DATE: 19920224
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandel, Saralynn
; REGISTRATION NUMBER: 31,853
; REFERENCE/DOCKET NUMBER: 8772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 796-4000
; TELEFAX: (818) 795-6321
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 662 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Oryctolagus cuniculus
; US-07-841-651-4

Query Match 10.4%; Score 308.5; DB 1; Length 662;
Best Local Similarity 23.4%; Pred. No. 1.9e-21;
Matches 154; Conservative 110; Mismatches 238; Indels 155; Gaps 26;
OY 11 IIVFYLLILVGIWAWTKNSGSAERSEAIIVGGRDGLLVGGFTMTATWVGGYING 70

Db 32 IVIYLVNVAWGLWAMEST-NRGTV---GGFFLAGRSMVWVPIGASLFAFNISGSHFVG 86
QY 71 TAEAVVPGYGLAWAQAPIGYS-----LSLILGLFFFAKPMRSKGYVTMLDPPQIY-GK 124
Db 87 LA-----CTGAASGATGCFENALMVVGLVFWVPIYIRA-GVVTPEYLOKRGKG 139
QY 125 RUGGLFTPALGEMFW---AAAFSALGAT-ISVIIDVDMHISVIAIATLYTVGG 181
Db 140 RIQIYLSLSLLYIFTKISADIFS--GAIFIQLTGLDIYVAIILLVGLTYITGGL 197
QY 182 YSVAVTVQVLCPIFGVGLWISVPEALSHPAVADIGTAVHAKY-----Q 225
Db 198 AAVIYTDIQTAIMVGVSVILGFAPEHVG-----GYEAFTEKYMRAIPSOISYGNISIPQ 253
QY 226 KPWLTGVSSEVYSWLDLFLMLGLGPW-----QAYFORVLSSSA 267
Db 254 KYCTPREDAFHI-----FRDAITGDPWPGVLFVFGMSILTLWVWCVDQVIVQRCLSAKNL 307
QY 268 TYAOVLSFLAARPGCLVMAIPAILIGAIGASTDNTQAYGLDP-----KTTEEADMLP 321
Db 308 SHVKGACILGGLKVMVPMFLVMMGMVSRILYTDKVCVVPSECECYCGTRVGCTNIAPF 367
QY 322 IVLOVLCPIYISFFGLGAVSAVMSADSSILSASSMFAINIYQLSFRONASDKETVWVM 381
Db 368 TLVVELMPLNGLRGLMVMASLMSLSLTSIFNSASTLFTMDIY-TKIRKASEKELMIAG 426
QY 382 RI-TVVFGASATAMALTKTVYG--LWVLSDLVIY--VIPQLLCVLFVKGNTYAV 436
Db 427 RLFLMLFLGISTAWPIVQSAQSGOLFQYISITVYLGPIIAAVELLAIFWKRVRNPGAF 486
QY 437 AGVYSGFLRI-----TG-----GPEYLYLQPLIFPGYPPDDNGIY 473
Db 487 WGLVGLGLIGISRMITEFAYGTGSCMEPSNCTIICGVHLYFAILF-----534
QY 474 NOKFPFKTAMVTSPLTNICISYLAKEYLFESGTLPPKLDVFAVVA-RHSEENMDKTIIV 532
Db 535 -----VISITVWVSLFTKPI-----PDVHLVRLCWSLRNSKE-----568
QY 533 KNEIKLD--ELAVKPKQSTLSTFTNKCAF-----LDVDSPESSGSGT 577
Db 569 --ERIDLDAGEDIEQAEATDEVPKKGKGFRRAYDLFGGLDQDQKPKMKEE 623

RESULT 2

US-07-841-651-2
; Sequence 2, Application US/07841651
; Patent No. 5410031

GENERAL INFORMATION:

; APPLICANT: Pajor, Ana M
; TITLE OF INVENTION: Cloning and Functional Expression of a
; TITLE OF INVENTION: Mammalian Na+/Nucleoside Cotransporter: A Member of the
; TITLE OF INVENTION: SGLT Family
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sheldon & Mak
; CITY: Pasadena
; STATE: California
; COUNTRY: USA
; ZIP: 91101

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,651
; FILING DATE: 19920224
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandel, Saralynn

; REGISTRATION NUMBER: 31,853
; REFERENCE/DOCKET NUMBER: 8772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 796-4000
; TELEFAX: (818) 795-6321
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 672 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-841-651-2

Query Match 10.0%; Score 298; DB 1: Length 672;

Best Local Similarity 25.0%; Pred. No. 2.1e-20;

Matches 153; Conservative 89; Mismatches 232; Indels 138; Gaps 25;

QY 9 IAIIVYLLILLVGINAAWRTKNSGSAERSEAIIVGGRDICLLVGGFTMTATVGGGY 67
Db 26 IAVIAAYFLLVIGVGLSMCRT-NRGTV---GGYFLAGRSMVWVPIGASLFAFNISGSH 80
QY 68 INGTAEAVYVPCYGLAWAQAPIGYSLS-----LILGLFFFAKPMRSKGYVTMLDPPQIY 123
Db 81 FVGLA-----GTGAANGAVAGFEWNAFVLLGLWLPFAPVLTAGVITM----PQYLR 130
QY 124 KRMG-----LLFTPALMGEMFWAAAIFF--SALGATISVIIDVDMHISVIA 169
Db 131 KRGGRIRLYLSVLSFLYIFTKISVDMFSGAVFIQALGNI-----YASVIAL 182
QY 170 LIATLYLVGLSVAYTDVVQVLCPIYISFFGLGAVSAVMSADSSILSASSMFAINIYQLSFRON 372
Db 183 GITWVTVTGLAALMYTDVQTEVIAAGAFILTYAGAPHEVG---GYSGFLDKYMGAMT 238
QY 225 -----OKPWLGTVDSEVYSWLDLFLLL---MLGSIW-----QAYF 258
Db 239 SLVSEDPAGVNISSSCYRPRPDSIHLLRDPVTDGLPMPALLGLTIVSGWYWCSDQIV 298
QY 259 QRVLSSSATYAOVLSFLAARPGCLVMAIPAILIGAIGASTDNTQAYGLPDKT---TE 314
Db 299 QRLAGRNLTIRKAGCILGKYLKTPMFLVMMGMVSRILYTDKVCVVPSECECYCGTRVGCTNIAPF 367
QY 315 E--ADMITPLVQVLCPIYISFFGLGAVSAVMSADSSILSASSMFAINIYQLSFRON 372
Db 359 VGCNSNTPRLVVKLMPNGLRGLMVAALMSLSLTSIFNSASTLFTMDIYTL--RPR 416
QY 373 SKKEIVWVRITVFEFGASATAMALTKTVYG---LWVLSDLVIY--IPQLLCVLFV 427
Db 417 GEGELLVGLWVVFIVAVSVANLPVVOAAGGOLFQYISVSSYLAPPVSAVVALFV 476
QY 428 KGTNTYGAAGVYVSGFLRTGGEPIYLYLQPLIFPGYPPDDNGIYNQKFPKTLAMY-- 485
Db 477 PRVNEKGAFWGLTGLLMLGLARLP-----EFSFGTGCVRP 513
QY 486 ---TSPLTNICISYLAKEYLFESGTLPPKLDVFAVVA-RHSEENMDKTI 530
Db 514 SACPFLCRVHLYFAILVFFCSGLLIITVSLCTAPIRKHHLRFLSLRHSKE-----567
QY 531 LVKNENIKLDEL 542
Db 568 --EREDLDDEL 577

RESULT 3

US-07-841-651-3
; Sequence 3, Application US/07841651
; Patent No. 5410031

GENERAL INFORMATION:

; APPLICANT: Pajor, Ana M
; APPLICANT: Wright, Ernest M
; TITLE OF INVENTION: Cloning and Functional Expression of a
; TITLE OF INVENTION: Mammalian Na+/Nucleoside Cotransporter: A Member of the
; TITLE OF INVENTION: SGLT Family
; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sheldon & Mak
 ; STREET: 225 South Lake Avenue, Ninth Floor
 ; CITY: Pasadena
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 91101
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/841.651
 ; FILING DATE: 19920224
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mandel, SaraLynn
 ; REGISTRATION NUMBER: 31,853
 ; REFERENCE/DOCKET NUMBER: 8772
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (818) 796-4000
 ; TELEFAX: (818) 795-6321
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 672 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Oryctolagus cuniculus
 ; US-07-841-651-3

Query Match 10.0%; Score 298; DB 1; Length 672;
 Best Local Similarity 25.0%; Pred. No. 2.1e-20;
 Matches 153; Conservative 89; Mismatches 232; Indels 138; Gaps 25;
 QY 9 IATVVFLLLVGTAWRTKNSGSAERSEAIIVGGRDIGLLVGGFTMTATWVGGY 67
 Db 26 IAVIAVFLVIGVGLSMCRT-NRGTV-----GGYFLAGRSVMWVPGASLFASNGSGH 80
 QY 68 INGTAEAVVPGYGLAWAQAIPIGYSL-----LILGLLFFAKPMRSKGYVTMLDPFQIYG 123
 Db 81 FVGLA-----GTGAAGLAVAGFEWALLFWLLGLWLPAPVILTAGVITM-----PQYLR 130
 QY 124 KRMGG-----LLFIPALMGEMFWAAIF--SALGATISVIIDVDMHISVLIISA 169
 Db 131 KRFGGHRIRLYSLVLSLFIYFTKISVDMFSGAVFIQOALGWN-----YASVIAL 182
 QY 170 LIATLYTLVGLYSVAVTVVQLFCIFVGLWISVPFALSHPAVADIGTAVHAKY----- 224
 Db 183 GITWVYVTGLAAMTDVQVTVIITAGAFILTYAFHEVG-----GYSGLFDKYMGMAT 238
 QY 225 -----QRPWLTGTVDSSEYVSLDSFLL-----MLGGIPW-----QAYF 258
 Db 239 SLTVSEDPAVGNISCCYRPRPSYHLLRDPVTDGLPWPALLGLITVSGWYSCDQIV 298
 QY 259 QRVLSSSATYAQVLSFLAAGCLVMAIPAILIAGASTDWNOTAYGLPDKT-----TE 314
 Db 299 QRCLAGRLTHIKAGCILCGYKLTLPMPFLWMPGMISILYDPDEVACVAPEVCRCVCGTE 358
 QY 315 E-ADMILPIVLOVLCPIVYISFFGLGAVSAVMSADSSILSASSMFARNIYQLSFRONA 372
 Db 359 VCCSNIAVPRVVKLMPNGLRGLMLAVMLAALMSSLSAIFNSSSTLFTMDIVTL--RPR 416
 QY 373 SDKEIVWMVNRITVVFVGASATAMALLTKTVYG---LWYLSDDLVIYV---IPQLLCVLEV 427
 Db 417 GEGELLVGRVWVIVAVSVAWLPVQAAGQGLFDYIOSVSSYLAPPVSAVFVVALFV 476
 QY 428 KGTNTYGAAGVSVGLFLRITGGEPLYLOPLIFYPGYPDDNGIYNOKFPEKTLAMV-- 485
 Db 477 PRVNEKGAFWGLIGLLMGLARLP-----EFSGTGSCVRP 513

QY 486 ---TSFLTNICISYLAHYLF-SG-----TLP-PKLDVDFDAVVA-RHSEENMDKTI 530
 Db 514 SACPAFLCRVHYLYPAIVLFFCSGLLIIVISLCTAPIPRKHLRLVFLSLRHSKE----- 567
 QY 531 LVKNENIKLDEL 542
 Db 568 --EREDLOADEL 577
 RESULT 4
 US-09-134-001C-4744
 ; Sequence 4744, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4744
 ; LENGTH: 518
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-4744

Query Match 8.8%; Score 262.5; DB 4; Length 518;
 Best Local Similarity 22.2%; Pred. No. 4.2e-17;
 Matches 126; Conservative 102; Mismatches 223; Indels 117; Gaps 25;
 QY 9 IATVVFLLLVGTAWRTKNSGSAERSEAIIVGGRDIGLLVGGFTMTATWVGGY 68
 Db 27 VMIIVFPIILLIIGFY-GYRQATGNLSE-----FMLGGRSIGPYITALSAGASMSGWMI 80
 QY 69 NGTAEAVVPGYGLAWAQAIPIGYSLILIGL--FFAKPMRSKY-----VTMLDPFQ 119
 Db 81 MGLPGSVYSGLSAIW-----ITIGLTGAYINVFVAPRLRYVTEIAGDAITLPDFK 134
 QY 120 QYQKRMGGLLFTPALMGEMFWAAIFSAI--GATISVIIDVDMHISVLIISALITATYT 176
 Db 135 NRLDDKKNIIKISGLIIVVEFTLYTHSGFVSGGKLFESAFGLNHAGLLIIVAIIVFYT 194
 QY 177 LVGGLYSVAVTVVQLFCIFVGLWISVPFALSHPAVADI-GFTAVHAKYQ-KPW----- 228
 Db 195 FFGYLAVSIIDFQGVIMLIAM-VMPVIV---ALLKNGWDTFHDIAQMKPTNLDLFR 249
 QY 229 ----LGTVDSEVYSWLDSEFLLLMLGGIPWQAYFORVLSSSSATYAQVLSFLAAGCLVM 284
 Db 250 GTTVLGIV---SLFSW-----GLGYFGQPHIIVREMSIKSHKLLPKARRLGISWM 296
 QY 285 AIPAILIAGASTDW-----NOTAYGLDPKTEADMLPIVLOVLCPIVYISFFGLGAV 340
 Db 297 AVG--LLGAIGVGLTGISFISERHKLDEPET-----LFIVMSQILPHPLVGGFLAAI 348
 QY 341 SAAVMSADSSILSASSMFARNIYQL---SFRONASDKREIVWMRITVVFVGASATAMAL 397
 Db 349 LAAINSTISSQLVTSSTLTFDYKLINGSDKASSHOKFVLIGRLSVLLVAIVAITIA- 407
 QY 398 LTKTVYGLWYLSDDLVIYV-----IFPQLLCVLPVKGTNTYVAGVYVSGFLFLRI 447
 Db 408 -----WHPNDTILNLVGNWAGFGAASPLVLSYWKDLTRAGALSGWAGAVVI 459
 QY 448 TGEPEYLYLOPLIFYPGYPDDNGIYNOKFPEKTLAMVTSFLTNIICISYLAHYLFSGTL 507
 Db 460 VW---ISWIKPLATINAF-----GMVE-----IIPCFIVSVLITIVSKL----TK 499
 QY 508 PKKLDVDFDAVVAHSEENMDKTIIVKNE 535

Db 500 KPD-----DYVI-----ENLNKVKHWKE 518

RESULT 5

US-08-595-553A-2

Sequence 2, Application US/08595553A

Patent No. 6391579

GENERAL INFORMATION:

APPLICANT: NANCY CARRASCO, ET AL.

TITLE OF INVENTION: THYROID SODIUM IODIDE SYMPORTER AND

TITLE OF INVENTION: NUCLEIC ACID ENCODING SAME

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN

STREET: 90 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 10016

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE

MEDIUM TYPE: DISKETTE

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/595,553A

FILING DATE: FEBRUARY 1, 1996

ATTORNEY/AGENT INFORMATION:

NAME: CRAIG J. ARNOLD

REGISTRATION NUMBER: 34,287

REFERENCE/DOCKET NUMBER: 96700/393

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 697-5995

TELEFAX: (212) 286-0854 or 286-0082

TELEX: TWX 710-581-4766

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 618

TYPE: AMINO ACID

TOPOLOGY: LINEAR

MOLECULE TYPE:

DESCRIPTION: PROTEIN

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: RAT

INDIVIDUAL ISOLATE: SODIUM/IODIDE SYMPORTER

US-08-595-553A-2

Query Match 7.3%; Score 218; DB 4; Length 618;

Best Local Similarity 20.8%; Pred. No. 1.3e-12;

Matches 125; Conservative 110; Mismatches 264; Indels 102; Gaps 23;

QY 13 VFYLLILL---VGIWAARWTKNSGSAERSEAIIVGGRIIGLVGGFTMTATWGGGVIN 69

Db 19 VFATMLIVSTGIGLVGLWGLARGGQSADD---FTGGRLAAPVGLSLAASFMSAVQVL 74

QY 70 GT-AEAVYVPGYGLAWAQAIPGIVSLILGLLFFAKPMRSKGYVTMLDPPQOYIGK--RM 126

Db 75 GVPAAEA---RYGLKFLWMCAGQLNLSLLTAFLEPIFYRLGLTSTYQYLELRFSAVRL 131

QY 127 GGLLFTPALMGEMFWAAAFSAIGATISVIIDVDMHSVIISALIAITLYTLVGLSVAY 186

Db 132 CGTL--QYLVATMLTYGIVIVAPALINQVTLGLDIWASLSTGICLYITVGGKAVVV 189

QY 187 TDVVOLFCEIPVGLWI-----SVFPALSHPAVADIGFTAVHAKYQKRWLGTV 232

Db 190 TDVQVVMVLGVFWLARGVILLGGPRNVLSLAQNHRSRLNLMDFDPPRSRYFTFWFIV 249

QY 233 DSSEVSWLDSFLLLMLGGIPQAYFORVLSSSSATVAQVLSFLAAGCLVMAIPALLIG 292

Db 250 GGLV--WLSMY-----GVN-QAQVRYVACHTEGAKLALLVNLGLFLIVASAACCG 300

QY 293 AIGASTDMNQATAYGLPDPKTT---EEADMILPIVQLV---CPVYISFFGLGAVSAVMS 346

Db 301 IV-----MFVYKDCDPLLTGRISAPDQVYMLPLLDIFEDLPQVCLF-LACAYSGLS 353

QY 347 SADSSILSASSMFARNIYQLSFRONASDKREIVVMRITVVFVGASATAMALLTKVYGLW 406

Db 354 TASTSINAAVAVTEDLIKPRM-PGLAPRKLVEISKGLSFIYGSACLTVAALSSLLGGV 412

QY 407 YLSSDLVYIVFIPQLCV---LFVKGTNTYGAAGVYVSGLELRI-----TGGEPEY 453

Db 413 LQGSFTVMGVISGPLGLAFTLGMLLPACHTPGVLSGLAAGLAVSLWVAVGATLYPPGEQT 472

QY 454 LYLQP-----LIFYPGYDDNGIYNQKFP-----FKTLAMV 485

Db 473 MGVLP TSAAGCTNDSVLLGPPGATNASNGIPSSGMDTGRPALADTFVAISLYLYGALGTL 532

QY 486 TSFLTNICISYLAKYLFESGTLPPKLDVDAVVARHSEENMDKTLVKNENIKLDELALV 545

Db 533 TTMCGALISYLTGPTKRS-SLGPGLLWMD--LARQT-----ASVAPKEDTATLEE-SLV 583

QY 546 K 546

Db 584 K 584

RESULT 6

US-09-134-001C-4510

Sequence 4510, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4510

LENGTH: 530

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4510

Query Match 6.9%; Score 206.5; DB 4; Length 530;

Best Local Similarity 22.7%; Pred. No. 1.3e-11;

Matches 117; Conservative 89; Mismatches 199; Indels 111; Gaps 26;

QY 9 IATIVFVLLILLVGIWAARWTKNSGSAERSEAIIVGGRIIGLVGGFTMTATWV----- 63

Db 4 ISILFIMVVGISFYAYLQSRKIKTSS--SDGYFMGNSL---TGFTVASTIIMTNLS 57

QY 64 -----GGGYING-----TAEAVYVPGYGLAWAQAIP---IGVSLSLILGLLFPK 104

Db 58 TEQIVQNGSGSYAQQGMEVMAWEVTAAVVVL---LAWFLPKYKLVKGVNISSELEURY 113

QY 105 PMRSKGYVTMLDPPQOYIKRMGGLLFI PALM-----GEMFWAAAFSAIGATISVII 157

Db 114 DFTFRFVSILFTYV-----VSFLPVLYSGSLVFNKMF---KVDELYGVSSSTAV 163

QY 158 DVDHSHVISALIAITLYTLVGLSVAYTDVVOLFCEIV-GLWISVPFALSHPAVADIG 216

Db 164 II---ISSII-GIIGIYLFIGLSLSAFSDSYGMALIIGGLAITI---LGLQLGDGN 216

QY 217 FTAVHAKYQK-----WLGTVDSSEVSWLDSFLLLMLGGI--PW-----QAYFORVLSSSS 266

Db 217 FLHGFDKIVQDTPPEKINGFGKVD-SDVWPWPTLFFGMFFNNLFFWCANQIVQKALAKN 275

QY 267 ATYAQ-----VLSFLAAGCLVMAIPAIL-IGAIGASTDWNQATAYGLPDPKTTTEADMILP 321

Db 80 QSAVALRVPSEIYRFTQYWFELRCCVFLGLLIPAHIFIPVYRLHLTSAYEYELERFNK 139
QY 105 PMRSKGYVTMLDPFQOI*GKRWGGLLIPALMGEMFMAAIFSAIGATISVIIDVDMHIS 164
Db 140 TVRVCGTFTPI--FOMVI--YMGVVLVAPSL-----ALNAVTFGLWLS 179
QY 165 VIISALIALTYLVGLYSVAYTDVQVLCFIVG-----LWISVPFAL 207
Db 180 VLALRVCTVYTAGLGLKAVITDVTQVFLVGLQALVIVGSAKVGGLGRVW---AVAS 236
QY 208 SHPAVADIGTAVHAKYQKPMWGLVDSSEVYSWLDSEVYVGLGIPQAVFORVLSSSA 267
Db 237 QHGRIS--GFELDPDFVRHTFTLAFGGV-----FMMLSLYGVN-OAQVORYLSRTE 287
QY 268 TYAQLVSFLAAF-----GCLVMAIPAIIIGAIGASTDWNOTAYCLPDKPTEEDM 318
Db 288 K-AAVLSYAVFPFQVSLVGC-----LIGLV-----MFAYQVEPMSIQQAQ 331
QY 319 IL-PIVLOYLCPVYISFFGL-GAVSAAMSSADSSILSASSMFARNIYQLSFR---QNAS 373
Db 332 APDQFVLYFVMDLLKGLPLGLFACIACISGSLTSSAFNSLATVMTMEDIRPWFPEFS 391
QY 374 DKEIYVWVRITVYFVGASATAMALLTKTVYGLWVLSDDL-----VYIVIFPQ-----L 421
Db 392 EARAIMLSRGLAFGYLLCLGMA-----YISSQMGVPLQNAISIFGMVGGPLGL 441
QY 422 LCV-LFVKGNTYTGAVAGVYVSGFLF 445
Db 442 FCLGNFFPCANPPGAVVGLLAGLV 466

RESULT 8
US-09-134-001C-4309
; Sequence 4309, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4309
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4309

Query Match 4.6%; Score 138; DB 4; Length 482;
Best Local Similarity 21.9%; Pred. No. 5.7e-05;
Matches 102; Conservative 78; Mismatches 176; Indels 110; Gaps 24;

QY 1 MAPHVEGLTAVFYLLILVGIWAARFKNSGAERSEAIIVGGRDGLLVGGFTMTA 60
Db 49 MGHAGGL-AIIIGWLITAGMISLALVFQN--LTNERSD-----LGGIYSYA 94
QY 61 TWVGGYINGTAEAVVPGVGLAWAQAPIGYSLSLILGLFFAKPMRSKGYVTMLDPFQ 120
Db 95 QAGFGDFI-----GFASW-----GYWFSAGLVNAYATLLMS-----IGNFFP 134
QY 121 IYKRMGGGLFIPALMGE-MFWAAIFSAIGATISVIIDVDMHISVIISALITLYLVG 179
Db 135 IF---KGNFTFPIIVASILLWSVHFLKGVETAALINSIVITIKLIPILLICMIVA 191
QY 180 -----GLYSVAY-----TDVOLFCEIVGLWISVFPALSHPAVD 214
Db 192 FNENTFRIGFGMDGYSLSLHFHANTMSQVKSTMLVTW-VFIGIEGAVVFSGRANKKD 250

Db 276 LKESQGAIVLSLFKVGGLITVPCWAFNPFNGSIDKSDNAY-----P 320
QY 322 IVLOYLCPVYISFFGL-GAV-SAAVMSADSSILSASSMFARNIYQLSFRONASDKEIVW 379
Db 321 ALVTSVLPWA--FGLGAVIFGAILSSVFGSLNSTITLTDFYKPIFGKNSDKHIAH 378
QY 380 VMRITVVEFGASATAMA---LLTKTVYGLWVLSDDLVIYIPQLLCVLFVKGNTNYGA 435
Db 379 VGHATVVGIVVVALAPVSLPSPGLYAVVQFNGVYSPVLAALILVAFPSKRTSLKGA 438
QY 436 -----VAGVVSGLFLRITGGEPLYLOPLIF 462
Db 439 KVTFLTHIILYAIISVFTEIN----YLTFTSVLIF 470

RESULT 7
US-09-014-969-11
; Sequence 11, Application US/09014969
; Patent No. 5965397
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/014,969
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-014-969-11

Query Match 6.7%; Score 200; DB 2; Length 635;
Best Local Similarity 23.8%; Pred. No. 7.5e-11;
Matches 120; Conservative 70; Mismatches 191; Indels 124; Gaps 23;

QY 3 FVHEGLTAVFYLLILVGIWAARFKNSGAERSEAIIVGGRDGLLVGGFTMTATW 62
Db 24 FSIIDVVFVLLVLSLAIGLYHACR----GWGRHTVGGELLMADRMGCLPVALSLAT 79
QY 63 VGGYINGTAEAVVPG-----YGL-----AKAQPIGYSLSLILG-----GLFFAK 104

Db 256 VSLGGLVGVVFAELLSLVTRFTRHVRVIEPGFVFIISLYL-----TSEMLS-LS 306
QY 243 SFLLMLGGIPWQAYFORVLSSSSATVAQVLSFLAAGC-----LVMAIPA1----- 289
Db 307 SILAITECGICCCOKYKAVANISQSAATTVRTMTKMLASGAETIIFMFLGISAVDPLIWTWN 366
QY 290 -----LITGAIGASTD-WNQAYGLPDPKPTTEADMLPVLVLOLCPVVISFFG 336
Db 367 TAFVRLTLLFVSFRFAGVGLVQVLTWLLNRYK-----VQLELDQVQVMSYGG 412
QY 337 L-GAVSAVMSADSSILSASSMF 359
Db 413 LRGAVAFALVALLDGNKVKKEKLF 436

RESULT 13

US-09-068-195-24
; Sequence 24, Application US/09068195B
; Patent No. 6140078

GENERAL INFORMATION:

; APPLICANT: Sanders, Jan W.
; APPLICANT: Ledebauer, Adrianus M.
; APPLICANT: Venema, Gerard
; TITLE OF INVENTION: Salt-Inducible Promoter Derivable from a Lactic Acid
; TITLE OF INVENTION: Bacterium, and its Use in a Lactic Acid Bacterium for
; TITLE OF INVENTION: Production of Desired Protein
; FILE REFERENCE: Sanders-60113/025227
; CURRENT APPLICATION NUMBER: US/09/068,195B
; EARLIER FILING DATE: 1998-07-29
; EARLIER APPLICATION NUMBER: PCT/EP97/04755
; EARLIER FILING DATE: 1997-08-20
; EARLIER APPLICATION NUMBER: EP 97200744/7
; EARLIER FILING DATE: 1997-03-13
; EARLIER APPLICATION NUMBER: EP 96202444/4
; EARLIER FILING DATE: 1996-09-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 503
; TYPE: PRT
; ORGANISM: L. lactis MGL363
US-09-068-195-24

Query Match 3.8%; Score 113.5; DB 4; Length 503;
Best Local Similarity 20.9%; Pred. No. 0.015;
Matches 119; Conservative 83; Mismatches 207; Indels 161; Gaps 28;

52 LVGGFTMTATWGGGYINGTAEAVVPGYGLAWAQAPIGYLSLILGGLFFAKPMRSKGY 111
8 LFGFFALTASMLVTVYEPT-----FATSKLHLVFFLLGGLLWFLPV----- 50
QY 112 VTMLDPFOOYIGKRMGGL-FIPALMGEMFNAATFSAIGATISVIIDVDMHISVIISAL 170
Db 51 ALCAAEAMATVGEWKNIGFVWSQTLGRFGAIF-----FQWFOITVGF 96
QY 171 IATLYTLVGLG-----YSVAYTD-----VVOLFVGLVWISVPPFALSHPAVADIGTAVH 221
Db 97 VTIYFIFLGALSYVNLQALNTDPLIFIGLIIIFWGLTF-----QLGQTORT 145
QY 222 AKYQKPW-LGTVDSSVYSWLDLFLMLGG-----IPWQAYFORVLSSSSATVAQVLSF 275
Db 146 AKLVKAGFVGVIVPSVILFGLAA--AYFIGNFIPIINSHAFVDPDSQVSTLVVVSF 203
QY 276 LAAF-----GCLVMAIPAILGATGASTDMQNTAYGLPDPKTT 313
Db 204 ILAYMGVEASASHINELNENKRNYPPLAMILLVILAILDAIGGFS-----VAAVIPQKELS 259
QY 314 EADMI---LPVILQY-----LCPVVISFFGLGAVSAVMSADSSILSASSMFA- 360
Db 260 LSAVGIOTFOTLILHFNHHLGLWVKVIALMIAFGVMGEVSWVVG-----PSRGMPAA 312

QY 361 --RNIYQLSFRONASDK---EIVVMVRITVVFEGASAT-----AMALLTKTVYG 404
Db 313 AQGRLLPKELRKTINTEVPVPLVMIQIIVTLGAVLTFGGGNNLSFLVAISLTVVYL 372
QY 405 LMYLSSDLVYVIFPOLLCVLFVKGTN---TYGAVAGYVSGFLRITGGEPYLYLQPLIF 461
Db 373 VGYL---LFFIYV-----VLIYKQNLKRTYN-VPGKIIG--KTIAGIGFL-LSIFAL 420
QY 462 YPGYPPDDNGIYNOKPFKTLAMVTSFLTNICISYLAAYL-----FESGTL 507
Db 421 FIFSVPPASIAKNETHYOMILLI-SFVTAIPLFFIYELHDKKHGDTIEEPTHFKAQDV 479
QY 508 PPKLDVFDVAVARHSEENMDKTLIVKNENI 537
Db 480 NPA--IYPAARGEHH-----IIRKEEHI 500

RESULT 14

US-08-295-814E-11
; Sequence 11, Application US/08295814E
; Patent No. 5658786

; GENERAL INFORMATION:
; APPLICANT: Smith, Kelli E.
; APPLICANT: Borden, Laurence A.
; APPLICANT: Hartig, Paul R.
; APPLICANT: Weinschank, Richard L.
; TITLE OF INVENTION: DNA ENCODING TAURINE AND GABA
; TITLE OF INVENTION: TRANSPORTERS AND USES THEREOF
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295,814E
; FILING DATE: DECEMBER 19, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40558-B-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 599 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-295-814E-11

Query Match 3.7%; Score 110; DB 1; Length 599;
Best Local Similarity 17.8%; Pred. No. 0.044;
Matches 92; Conservative 59; Mismatches 146; Indels 220; Gaps 24;

QY 29 TKNSSGAE-----ERSEAIIVGGRDGLLV-----GGFTMTATWGGGYINGTAEAVV 77
Db 3 TNSKVADQCIISTEVSEAPVADSKPTLVVYVQKKAGDLPDRDTWK----- 49
QY 78 PGYGLAWAQAPIGYLSLILGGLFFAKPMRSKGYVTMLDPFOOYIGKRMGGLFIPALMG 137
Db 50 ---REDFLMSCVGYAIG--LGNVW-----RFPYLCGNKGAGFLIPIYFLT 89

QY 138 EMFWAAIF---SALGATISVIIDVDMHISVIISALIAITLYTLVGGLYSVAYTDVVQLFC 194
Db 90 LIFAGVPLFLLCGLGQ-----YTSIGGL----- 113
QY 195 IFVGLWISVPFALSHPAVADIGFTAVHAKYOKPWLGTVDSSSEVYSWLDSEFLLMLGIPW 254
Db 114 ---GWKGLAPM-----FKGVGLAAAVLSF-----WLNIIYIIVI---ISW 146
QY 255 QAYFORVLSSSSSATYAQVLSFLAAGCLVMAIPAAILIGAIGASTDNNOTAYGLPDPKTE 314
Db 147 AIYY-----LYNSETTLTPWKQ-----CDNPWNTD 171
QY 315 EADMILPIVQLCPVYISFFGLGAVSAAMSSADSSILSASSMFARNIYQLSFRQNASD 374
Db 172 RC-----FSNYSLVNTTNTSA-----VVEFWERNHQMT---DGLD 205
QY 375 K--EIVWVMRITVVFPGASATAMALLTKTVYGLW-----YLSDDLVIYVIFPOLL 422
Db 206 KPGQIRWPLAITL-----AIAWVLV---YFCIRKGVGWTGKVYFSATPYI-----ML 251
QY 423 CVLFVKGTNTYGAVAG---YVSGFLRITGGEPLYLQPLIFYP-----GYYP 467
Db 252 IILFRGVTLPGAKEGILFYITPNFRKLSDESEWLDAAQTQIFFSYGLGSLIALGSGNS 311
QY 468 DDNGIYNOKFPFKTLAMVTSFLTNICISYLAKEYLFS 504
Db 312 FHNNVYRDS-----IIVCCINSCTSMEFAGVIFS 340

RESULT 15

US-08-240-783B-4
; Sequence 4, Application US/08240783B
; Patent No. 5756348
; GENERAL INFORMATION:
; APPLICANT: Smith, Kelli
; APPLICANT: Borden, Laurence A.
; APPLICANT: Branche, Theresa
; APPLICANT: Hartig, Paul R.
; APPLICANT: Weinschank, Richard L.
; TITLE OF INVENTION: DNA ENCODING A GLYCINE TRANSPORTER AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/240,783B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 1795/39875-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 599 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: N
; FRAGMENT TYPE: internal

; ORIGINAL SOURCE:
; ORGANISM: RAT GABA TRANSPORTER (GAT-1)
US-08-240-783B-4

Query Match 3.7%; Score 110; DB 1; Length 599;

Best Local Similarity 17.8%; Pred. No. 0.044;
Matches 92; Conservative 59; Mismatches 146; Indels 220; Gaps 24;

QY 29 TKNSSGSAE-----ERSEAIIVGGRDIGLLV-----GGFTMTATWVGGYINGTAEAVV 77
Db 3 TONSKVADGQISTEVSEAPVADSKPTLVVKVQKAGDLPDRDTWKG----- 49
QY 78 PGYGLAWAQAIPGYSLSLILGGFFFAKPMRSKGYVTMLDPPQQIYKRMGGLIFPALMG 137
Db 50 ---RFDLMSCVGYAIG--LGNVW-----RFPYLCGKNGGGAFLIPYFLT 89
QY 138 EMFWAAIF---SALGATISVIIDVDMHISVIISALIAITLYTLVGGLYSVAYTDVVQLFC 194
Db 90 LIFAGVPLFLLCGLGQ-----YTSIGGL----- 113
QY 195 IFVGLWISVPFALSHPAVADIGFTAVHAKYOKPWLGTVDSSSEVYSWLDSEFLLMLGIPW 254
Db 114 ---GWKGLAPM-----FKGVGLAAAVLSF-----WLNIIYIIVI---ISW 146
QY 255 QAYFORVLSSSSSATYAQVLSFLAAGCLVMAIPAAILIGAIGASTDNNOTAYGLPDPKTE 314
Db 147 AIYY-----LYNSETTLTPWKQ-----CDNPWNTD 171
QY 315 EADMILPIVQLCPVYISFFGLGAVSAAMSSADSSILSASSMFARNIYQLSFRQNASD 374
Db 172 RC-----FSNYSLVNTTNTSA-----VVEFWERNHQMT---DGLD 205
QY 375 K--EIVWVMRITVVFPGASATAMALLTKTVYGLW-----YLSDDLVIYVIFPOLL 422
Db 206 KPGQIRWPLAITL-----AIAWVLV---YFCIRKGVGWTGKVYFSATPYI-----ML 251
QY 423 CVLFVKGTNTYGAVAG---YVSGFLRITGGEPLYLQPLIFYP-----GYYP 467
Db 252 IILFRGVTLPGAKEGILFYITPNFRKLSDESEWLDAAQTQIFFSYGLGSLIALGSGNS 311
QY 468 DDNGIYNOKFPFKTLAMVTSFLTNICISYLAKEYLFS 504
Db 312 FHNNVYRDS-----IIVCCINSCTSMEFAGVIFS 340

Search completed: March 31, 2003, 18:38:34

Job time : 43 secs



11
c